Integrating Genomic and Phenomic Breeding Selection Tools with Field Practices to Improve Seed Composition and Quality Traits in Soybean

William Monte Singer

Dissertation submitted to the faculty of the Virginia Polytechnic Institute and State University in partial fulfillment of the requirements for the degree of

Doctor of Philosophy
In
Crop and Soil Environmental Sciences

Bo Zhang, Chair
David Holshouser
Haibo Huang
Rouf Mian
Scott Raubenstine

October 28th, 2021
Blacksburg, Virginia

Keywords: soybean, methionine, amino acids, fatty acids, genome-wide association, genomic selection, phenomic selection

Integrating genomic and phenomic breeding selection tools with field practices to improve seed composition and quality traits in soybean

William Monte Singer

ACADEMIC ABSTRACT

Despite soybean's widespread recognition as a versatile and valuable crop due to many end-use purposes, breeders seek to develop varieties with improved nutritional and functional components that capture added-value for producers. Additionally, producers seek to maximize profits by utilizing field practices to augment crop value. Therefore, this dissertation had two main objectives of maximizing soybean value: 1) to evaluate accelerated selection methods by soybean breeders for methionine content and test weight, and 2) to identify sulfur fertilization impact on soybean seed composition including amino and fatty acid profiles. First, a genome-wide association study (GWAS) analyzed genomic influence on proteinogenic methionine in soybean seeds which identified 23 single nucleotide polymorphisms (SNPs). Utilizing a SNPs subset identified by GWAS, genomic selection (GS) exhibited average prediction accuracies ranging from 0.41-0.62. Secondly, a novel phenomic selection (PS) method using near-infrared reflectance spectroscopy (NIRS) was evaluated for predictive ability of soybean test weight. PS cross-validations exhibited average predictive accuracies of 0.75, 0.59, and 0.16 when incorporating all environments, between locations, and between years, respectively. Finally, sulfur fertilizer rates and sources were assessed across two years and six locations in relation to seed composition. Notably, ammonium sulfate (AMS) was found to have a significant impact ($P < 0.05$) on methionine content in soybean seed. These outcomes will have positive impacts on plant breeding and soybean production for seed composition and quality traits using contemporary breeding and fertilization.

Integrating genomic and phenomic breeding selection tools with field practices to improve seed composition and quality traits in soybean

William Monte Singer

PUBLIC ABSTRACT

Despite soybean's widespread recognition as a versatile and valuable crop due to a myriad of end-use purposes, breeders seek to develop varieties with improved nutritional and functional components that captured value for producers. Additionally, producers seek to maximize their profits by utilizing field practices that increase crop value. Therefore, this dissertation had two main objectives of maximizing soybean value: 1) to evaluate accelerated selection methods by soybean breeders for methionine content and test weight, and 2) to identify sulfur fertilization impact on soybean seed protein and oil composition. The overall objective was to create a comprehensive toolset for soybean breeders to develop Mid-Atlantic soybean varieties with improved seed composition traits and to determine fertilization impacts for use by producers. Genetic controls for protein-bound methionine in soybean seed were identified and could be used for variety development. Additionally, a new prediction method that uses light reflectance to represent genetic information and environmental effects was shown to have high accuracy for soybean test weight. It was also found that sulfur fertilizer with high availability in the soil positively impacted methionine content. These outcomes will have positive impacts on plant breeding and soybean production for seed composition and quality traits using contemporary breeding and fertilization.

Acknowledgements

I would like to thank Dr. Bo Zhang, my graduate advisor, for all of the support and mentorship she provided during my time as a graduate student. She encouraged the challenges I set for myself and offered advice when those challenges seemed insurmountable. She constantly pushed me to be the best researcher possible, and I would not have achieved this level of success without her guidance. I would also like to thank my committee members, Dr. Haibo Huang, Dr. David Holshouser, Dr. Rouf Mian, and Scott Raubenstine, for their continual support and expertise throughout all of my projects.

Research is done best through collaboration and teamwork, so everyone who helped during these projects also deserve thanks. Specifically, I would like to thank current and past members of the soybean breeding program at Virginia Tech: Luciana Rosso, Muliang Peng, Nilanka Lord, Zachary Shea, Xiaoying Li, Jessica Wilbur, Patrick Bewick, Xingbo Wu, Zhibo Wang, Mathew Colson, Qian Zhu, Justin Polk, and Elizabeth Prenger; undergraduate researchers: Lauren Seeley and Mackenzie Woolls; members of the Eastern Virginia Agriculture Research and Extension Center: Joseph Oakes, Lin Barrack, Michelle Lee, and Mark Vaughn; as well as Chao Shang.

Finally, I would like to thank my family for all of their support during my time as a graduate student. I would like to thank my parents, Pete and Becky Singer, and my brother, Michael Singer for believing in me since I was little as well as my family-in-law who welcomed me with open arms. I especially thank my wife and partner, Margaret Nagai-Singer, for her unwavering support through it all.

Attributions

Many researchers provided expertise to the projects described in this dissertation. Their names, departments, institutions, and role description are provided for each chapter.

Chapter 1: Soybean Amino Acids in Health, Genetics, and Evaluation

William M. Singer, Ph.D. Candidate, School of Plant and Environmental Sciences, Virginia Tech. Mr. Singer explored and reviewed a comprehensive collection of previous research that he summarized into a textbook chapter.

Bo Zhang, Assistant Professor, School of Plant and Environmental Sciences, Virginia Tech. Dr. Zhang reviewed and edited the final manuscript.

Rouf Mian, USDA Professor, Department of Crop and Soil Sciences, North Carolina State University. Dr. Mian reviewed and edited the final manuscript.

Haibo Huang, Associate Professor, Department of Food Science and Technology, Virginia Tech. Dr. Huang reviewed and edited the final manuscript.

Chapter 2: Genome-wide Association Study and Genomic Selection for Proteinogenic Methionine in Soybean Seeds

William M. Singer, Ph.D. Candidate, School of Plant and Environmental Sciences, Virginia Tech. Mr. Singer was responsible for performing all experiments, managing field plots, data analysis, and writing the manuscript.

Zachary Shea, Ph.D. Student, School of Plant and Environmental Sciences, Virginia Tech. Mr. Shea assisted with germplasm management and harvest. He also edited the manuscript.

Dajun Yu, Ph.D. Candidate, Department of Food Science and Technology, Virginia Tech. Ms. Yu provided liquid chromatography assistance.

Haibo Huang, Associate Professor, Department of Food Science and Technology, Virginia Tech. Dr. Huang assisted with experimental methods and edited the manuscript.

Rouf Mian, USDA Professor, Department of Crop and Soil Sciences, North Carolina State University. Dr. Mian assisted with experimental objectives and edited the manuscript.

Chao Shang, Senior Research Associate, School of Plant and Environmental Sciences, Virginia Tech. Dr. Shang provided guidance on methionine hydrolysis and quantification.

Maria L. Rosso, Faculty Research Associate, School of Plant and Environmental Sciences, Virginia Tech. Dr. Rosso provided guidance for methionine quantification. She also edited the manuscript.

Qijian Song, Research Geneticist, Soybean Genomics and Improvement Laboratory Beltsville Agricultural Research Center, USDA-ARS. Dr. Song assisted in selecting the research materials. He also edited the manuscript.

Bo Zhang, Assistant Professor, School of Plant and Environmental Sciences, Virginia Tech. Dr. Zhang assisted with experimental objectives and edited the manuscript.

Chapter 3: Comparison of Genomic and Phenomic Prediction Accuracy for Soybean Test Weight

William M. Singer, Ph.D. Candidate, School of Plant and Environmental Sciences, Virginia Tech. Mr. Singer was responsible for performing all experiments, managing field plots, data analysis, and writing the manuscript.

Zachary Shea, Ph.D. Student, School of Plant and Environmental Sciences, Virginia Tech. Mr. Shea assisted with germplasm management, harvest, and test weight measurement. He also edited the manuscript.

Bo Zhang, Assistant Professor, School of Plant and Environmental Sciences, Virginia Tech. Dr. Zhang assisted with experimental objectives and edited the manuscript.

Chapter 4: Soybean Amino Acid and Fatty Acid Response to Sulfur Fertilization

William M. Singer, Ph.D. Candidate, School of Plant and Environmental Sciences, Virginia Tech. Mr. Singer was responsible for trait quantification, data analysis, and writing the manuscript.

Keren Brooks, Ph.D. Student, School of Plant and Environmental Sciences, Virginia Tech. Mrs. Brooks assisted in experimental design and managed field locations. She also edited the manuscript.

Mark Reiter, Associate Professor, Virginia Tech Eastern Shore Agricultural Research and Extension Center, Virginia Tech. Dr. Reiter assisted with experimental objectives and edited the manuscript.

David L. Holshouser, Associate Professor, Virginia Tech Tidewater Agricultural Research and Extension Center, Virginia Tech. Dr. Holshouser assisted with result interpretation and edited the manuscript.

Bo Zhang, Assistant Professor, School of Plant and Environmental Sciences, Virginia Tech. Dr. Zhang assisted with experimental objectives and edited the manuscript.

Table of Contents

List of Tables

List of Figures

Introduction

Soybean [*Glycine max (L.) Merril*] is a member of the *Fabaceae* family and an annual leguminous crop grown globally across diverse environments with assorted management practices. Originating in East Asia, soybean has been used for human food and livestock feed for centuries, and its global production has increased 350% since 1987 (Soy Meal Info Center, 2018). This tremendous production growth is directly linked to soybean seed composition and the international demand for protein and oil sources. The soybean seed consists of roughly 40% protein and 20% oil which are comprised of their respective monomers, amino and fatty acids (Wilson, 2004). Modern plant breeders, especially in the United States, have been successful in optimizing the productivity of soybean while also adding value by augmenting seed composition.

Continuous genetic improvement of soybean and the development of new varieties are critical to increasing crop value and supporting global food systems. Therefore, in this dissertation potential breeding tools and management decisions were explored to maximize seed composition and quality. The first chapter contains a comprehensive review of amino acids in soybean seeds. The first study seeks to identify a genetic basis for proteinogenic methionine in soybean seeds by using a genome-wide association study and genomic prediction. The second study compares phenomic and genomic prediction for test weight, an indicator of seed quality, in soybean. The final study analyzes environmental impacts on amino and fatty acids in soybean seeds through a sulfur fertilizer trial. In combination, results from these studies can assist breeders in developing soybean varieties with improved seed compositions and expand producer knowledge for management decisions.

References

Soy Meal Info Center. 2018. World Soybean Production. https://www.soymeal.org/soy-meal-articles/world-soybean-production/ (accessed 20 September 2021)

Wilson, R.F. 2004. Seed Composition. Soybeans: Improvement, Production, and Uses. 3rd ed. ASA, CSSA, SSSA, Madison, WI. p. 621–669

Chapter 1: Soybean Amino Acids in Health, Genetics, and Evaluation

William M. Singer[1], Bo Zhang[1], Rouf Mian[2] and Haibo Huang[3]

[1]School of Plant and Environmental Sciences, Virginia Tech

[2] Soybean & Nitrogen Fixation Unit, USDA-ARS

[3]Department of Food Science and Technology, Virginia Tech

This chapter was published by IntechOpen on October 7th, 2019:

Singer, W.M., Zhang, B., Mian, M.A.R., and Huang, H. (2019). Soybean Amino Acids in Health, Genetics, and Evaluation. In: *Soybean for Human Consumption and Animal Feed*. IntechOpen. DOI:10.5772/intechopen.89497

Abstract

Soybean is an important source of protein and amino acids for humans and livestock because of its well-balanced amino acid profile. This chapter outlines the strengths and weaknesses of soybean as a complete amino acid source as well as the relative importance of individual amino acids. Special attention is paid to the sulfur-containing amino acids, methionine and cysteine. Breeding and genetic engineering efforts are summarized to highlight previous accomplishments in amino acid improvement and potential avenues for future research. Agronomic properties and processing methods that affect amino acid levels in soybean food and feed are also explained. A brief introduction into current amino acid evaluation techniques is provided. By understanding the complexities of amino acids in soybean, protein quality for humans and livestock can be maximized.

Introduction

Soybean is one of the world's most economically and nutritionally important crops. In 2018, soybean was 61% of international oilseed production with 397.9 tons harvested worldwide (The American Soybean Association, 2019a). The United States and Brazil were the largest producers at 4,545 and 4,299 million bushels, respectively, with China being the largest importer of U.S. whole soybeans valued over $3 billion (The American Soybean Association, 2019bc). Soybean products, namely meal and oil, are popular in a myriad of industries for their versatility and utility. Soybean oil provides the most versatility with uses in fuel, solvents, candles, cosmetics, construction, and foam. However, soybean meal is the driving factor for 70% of the plant's value with 97% of all U.S. soybean meal being used for animal feed [United Soybean Board(b)]. As such, an enormous portion of soybean's importance lies with its nutritional capabilities for livestock and humans.

Nutritionally speaking, soybeans are a highly valued protein source. Proteins are a crucial macromolecule needed in the diets of human and livestock. However, the true significance of soybean protein is due to its well-balanced amino acid profile that aligns with dietary needs of humans and animals (Osborne and Mendel, 1993). Amino acids are the functional subunits of proteins that, when linked together in different orders, generate the variety of proteins critical to life. Amino acids are also important intermediates for many biosynthesis pathways (Herrmann and Somerville, 1983). Deficiencies in single or multiple amino acids can negatively impact an individual's growth and development (Berry et al., 1962; Wade, 1985). Intriguingly, an excess of certain amino acids has also been shown to worsen feed intake, nitrogen efficiency, and growth rate in livestock

(Waldroup et al., 1976; Han et al., 1992; Boisen et al., 2000). The importance of amino acid levels on human health has also been well documented (Jez and Fukagawa, 2008; D'Mello, 2012).

Amino acids are characterized by having amine ($-NH_2$) and carboxyl (-COOH) functional groups as well as a "R-group" that is unique to each amino acid (Weaver, 2008). Amino acids are abundant in both proteinogenic (protein-incorporated) and non-proteinogenic forms (Wagner and Musso, 1983). The 20 common, proteinogenic amino acids are generally the focus of research in soybeans as they are the defining nutritive feature. Of those 20, nine amino acids are essential for humans to consume. Livestock usually require these same amino acids from feed and might require others because of their biological systems (Buttery and D'Mello, 1994). Soybeans contain some level of all nine essential amino acids which creates a suitable nutritional foundation for livestock feed and human food.

Essential Amino Acids

Essential amino acids are ones that living organisms are unable to biosynthesize themselves and must obtain from their food source (Wade, 1985; Buttery and D'Mello, 1994; D'Mello, 2012). Therefore, in this term, "essential" refers to the amino acid requirements in dietary ingredients. The nine standard essential amino acids for humans present in soybean are: histidine, isoleucine, leucine, lysine, methionine, phenylalanine, threonine, tryptophan, and valine (Kuiken and Lyman, 1949). Arginine is regularly considered an essential amino acid for fish, poultry and sometimes swine due to absent or deficient urea cycles (Buttery and D'Mello, 1994; Boisen et al., 2000). Poultry and reptiles also require dietary glycine because of differing waste excretion pathways

(Buttery and D'Mello, 1994). While crude protein content is normally recognized as the driving nutritional factor for soybean meal, these essential amino acids provide true utility.

It has long been recommended that protein quality is based upon essential amino acid content. However, for many reasons, animal feed and human food markets have only recently begun assessing accordingly [Osborne and Mendel, 1993; Pfarr et al., 2018; United Soybean Board(a)]. Equipment required for accurate amino acid measurement and the diversity of markets for amino acids makes it difficult for supply chain evaluators like elevator operators to appraise amino acid content on site. To some degree, the well-balanced soybean amino acid profile also devalues the need to measure individual amino acid levels. Since all essential amino acids are present, less attention is paid to deficient amino acids such as methionine and tryptophan (Kuiken and Lyman, 1949; Fernandez et al., 1994).

Deficiencies in soybean's essential amino acid profile has led to a large section of the livestock industry focusing on feed mixing and supplementation. Rationing with other feed sources such as cereal grain and synthetic amino acid augmentation can effectively resolve the issue. Although, this comes with economic and environmental problems. Supplementing amino acids adds costs to farmers. For example, the average cost for amino acids supplementation for dairy farmers is twenty cents per head per day (Drovers). Maximizing crude protein for a growth limiting factor also negatively impacts livestock nitrogen-use-efficiency and environmental nitrogen outputs (Berry et al., 1962; Meisinger, 2005). Synthetic amino acid production can produce hazardous environmental waste and synthetic methionine, the most limiting soybean amino acid for poultry, has

also been banned for organic poultry production (Fernandez et al., 1994; Willke, 2014). Movement towards sustainable agriculture will pressure the feed industry to alter how soybean meal is enhanced for essential amino acid livestock maximization. Furthermore, the increasing popularity of meat-less diets in humans will create new markets for soybean's well-balanced amino acid profile.

Non-essential Amino Acids

Non-essential amino acids should not be misconstrued as unimportant amino acids. Of the twenty proteinogenic amino acids, those considered non-essential are still necessary for living organisms. Healthy organisms are just able to biosynthesize them and are not from food and feed consumption. The eleven standard non-essential amino acids for humans found in soybean are: arginine, alanine, asparagine, aspartic acid, cysteine, glutamine, glutamic acid, glycine, proline, serine, and tyrosine (Rackis et al., 1961; Goldflus et al., 2006; Kita et al., 2010). As previously mentioned, the necessity of amino acids such as arginine and glycine can differ amongst species. Some non-essential amino acids are also affected by the presence and amounts of essential amino acids.

Cysteine not provided from food consumption is directly biosynthesized from methionine via trans-sulfuration (Kredich, 1983; Buttery and D'Mello, 1994; Fuller, 1994; Brosnan and Brosnan, 2006; Mato et al., 2012). Consequently, if cysteine is not provided in the diet, then enough methionine must be provided to compensate for both amino acid needs. For that reason, feed research for poultry occasionally measures methionine and cysteine jointly (Han et al., 1992; D'Mello, 1994; Jankowski et al., 2014). Tyrosine is also directly formed from phenylalanine via hydroxylation (Herrmann, 1983; Buttery and D'Mello, 1994; Fuller, 1994). Other amino acids like arginine, glycine,

and proline can be required from the diet when animals are young, old, sick, or otherwise deficient in body protein regulation. As human and livestock diets become more sustainably plant-based, it will become more important to evaluate non-essential amino acids, specifically the ones immediately affected by essential amino acids.

Proteinogenic Sulfur-containing Amino Acids

The two proteinogenic sulfur-containing amino acids, methionine and cysteine, are critical to evaluate soybean meal as food and feed. While present in soybean, methionine and cysteine levels are both inadequate for consumer needs (Berry et al., 1962; Fernandez et al., 1994; Goldflus et al., 2006). Similar to research determinations, the nutritional requirements for methionine and cysteine intake are often grouped together as overall sulfur consumption from protein. Adult humans are recommended to intake 910-1,120 mg methionine and cysteine (based on body weight) per day (Jez and Fukagawa, 2008). For livestock, sulfur amino acids recommendations can vary based on species, age, end-use, and diet formulation. The importance of dietary sulfur amino acids for livestock is greatly emphasized in literature, especially with soybean meal as base feed (Berry et al., 1962; Fernandez et al., 1994; Allee, 2003; Wu, 2014).

Methionine and cysteine are vital to biological functions because of the sulfur contained in their R-groups and versatility in macromolecule synthesis. Methionine is well-known for being the typical initiating amino acid for protein synthesis and has hydrophobic properties when incorporated into proteins (Ingenbleek and Kimura, 2013). These hydrophobic properties usually result in methionine incorporation within the core of proteins. However, certain proteins have surface-exposed methionine susceptible to oxidation that is associated with age-related disease (Levine et al., 1996; Moskovitz,

2005; Brosnan and Brosnan, 2006). S-adenosylmethionine, a methionine metabolism intermediate, is widely-used with functions in methylation as well as amine, methylene, and sulfur atom donation (Catoni, 1953; Fontecave et al., 2004; Brosnan and Brosnan, 2006). Methionine is especially important for poultry production as birds have exceedingly high sulfur amino acid requirements, and low methionine levels can negatively affect growth rate, carcass yield, fat content, and disease immunity (Bunchasak, 2009; Wu et al., 2012; Conde-Aguilera et al., 2013). Cysteine's ability to form disulfide bonds makes it incredibly important to tertiary protein structure and can occur with and without enzyme interactions (Jessop et al., 2004; Brosnan and Brosnan, 2006). Cysteine is involved with keratin and feather production in poultry and deficiencies have been correlated to poor breast muscle development (Wylie et al., 2001; Bonato et al., 2011). Swine also need higher amounts of cysteine as they age to compensate for body maintenance (Fuller, 1994; Lewis, 2003).

Breeding Efforts

Soybean has an extensive cultivated history dating back thousands of years to its country of origin, China. *Glycine max*, the contemporary species of soybean, was domesticated from the wild species *Glycine soja* and has been continually been improved through selective and molecular breeding (Hermann, 1962; Liu, 1997; Sleper and Poehlman, 2006). Once harvesting traits such as seed shatter and lodging were improved in the late 1930's to make soybean a competitive row crop, other cultivar improvements became a valuable research goal (Sleper and Poehlman, 2006). Current breeding programs tend to focus on traits such as yield, disease resistance, abiotic stress tolerance, and seed composition. Seed composition improvements include protein and oil content,

fatty acid levels, anti-nutritional factors, isoflavones, and amino acids profiles. Before 1972, there had been zero reported research for improvement of soybean amino acid profiles, rather with emphasis on overall protein content (Howell et al., 1972). Modern breeders are also inclined to concentrate efforts on protein content and consider amino acid levels an afterthought (Mahmoud et al., 2006). TN04-5321 is the only released germplasm in the United States that maintains yield and protein content while improving amino acid balance by increasing methionine and cysteine to levels recommended livestock needs (Panthee and Pantalone, 2006).

The major soybean storage proteins are 11S (glycinin) and 7S (conglycinin) and provide the bulk of amino acids with limited non-proteinogenic amino acids in seed (Wolf, 1969; Meinke et al., 1981; Takahashi et al., 2003). By increasing 11S and 7S quantity, more protein can ultimately be present in food and feed. Overemphasis on crude protein content can have negative ramifications on overall protein quality, specifically deficient amino acids. While an increase in protein content would theoretically entail an increase in amino acids including methionine and cysteine (Wilcox and Shibles, 2001), the opposite effect has been more notable (Paek et al., 1997; Thakur and Hurburgh, 2007). Molecular breeding techniques have recently improved the understanding of amino acid genomic regulation. Multiple quantitative trait loci (QTL) studies have been performed to identify genomic regions that control amino acids levels in soybean seed (Panthee et al., 2006b; a; Fallen et al., 2013; Warrington et al., 2015; Li et al., 2018). A myriad of QTL's was found to create amino acid phenotypic variation. Individual amino acids had reoccurring or proximal QTL's discovered such as Satt 518 (Panthee et al., 2006b), ss107913002 (Fallen et al., 2013), and BARC-048619 (Warrington et al., 2015)

for glycine and threonine. QTL's for methionine and cysteine were also discovered which could lead to valuable improvements for soybean livestock feed (Panthee et al., 2006b; a; Fallen et al., 2013; Warrington et al., 2015; Li et al., 2018). Other genomic studies such as genome-wide association studies (GWAS) and genetic diversity analyses would further improve genetic understanding.

Genetic Engineering

Genetic engineering experiments such as genetically modified organisms (GMO's) and gene editing are also promising avenues for improving amino acid profiles of soybeans. Compared to conventionally bred varieties, transgenic soybeans face additional adversity from registration requirements and public opinion. Transgenic efforts generally have one of three targets: magnifying biosynthesis genes, adjusting biosynthesis regulation, and modifying storage proteins. The earliest example would be a Brazil nut gene transfer in 1992. This successfully increased protein content and methionine biosynthesis, however a major food allergen was also transferred making commercialization impossible (Townsend et al., 1992). Expressing zein proteins from corn has also been shown to increase sulfur-containing amino acids levels in soybean (Kerr, 1996; Kim and Krishnan, 2019). Altering biosynthesis feedback regulation amplified both non-proteinogenic and proteogenic lysine by circumventing normal enzymatic pathways (Falco et al., 1995). Tryptophan in soybean also exhibited increased non-proteinogenic levels when a feedback-insensitive enzyme was transferred (Kita et al., 2010). While soybean is not deficient in lysine or tryptophan, corn is deficient in both. By increasing lysine and tryptophan concentrations, soybean becomes an even more useful feed additive to corn rations. Even with limited research on modifying overall amino

acids profiles in soybeans, modifying 11S and 7S storage proteins ratios (El-Shemy et al., 2007) or silencing their expression entirely (Schmidt et al., 2011) has displayed increased amino acids levels. Similarly, a study using irradiated mutant soybeans lacking storage proteins as breeding parents demonstrated increased non-proteinogenic amino acids contents (Takahashi et al., 2003). In addition, further research should be conducted to determine the bioavailability and digestibility of increased non-proteinogenic amino acids in soybean.

Agronomic Relations

Amino acids concentrations in soybean are not only affected by their genetic potential. Agronomic properties greatly impact the final levels of amino acids. Agronomy encompasses all aspects of crop production including environmental effects, climatic variables, and abiotic factors. Perhaps the most considered agronomic factor is soil nutrient availability. Insufficient soil nutrient levels of nitrogen, potassium, phosphorous, sulfur, calcium, and magnesium create poor amino acid profiles in soybean plants (Haghiri, 1966). Increased phosphorous rates have been shown to increase the percentage of methionine and tryptophan in seed but had no effect on protein content percentage (Kapoora and Gupta, 1977). Applications of sulfur, phosphorous, and nitrogen (individual and combined) produced a variety of different methionine and cysteine seed concentrations (Arora and Luthra, 1971). Sulfur deficiencies were also shown to inhibit the production of 11S proteins while almost eliminating methionine and cysteine in 7S proteins (Gayler and Sykes, 1985). It is becoming more popular to also apply biological substances such as amino acids to plants through foliar and seed application. Amino acid uptake by soybean and wheat have been proven, and improved soybean growth rates and

antioxidant effects have also occurred (Gioseffi et al., 2012; Teixeira et al., 2017). Further research should be conducted to determine if biofortification solutions are possible through amino acid application.

Amino acid variation has also been shown to occur across environments (Goldflus et al., 2006; Thakur and Hurburgh, 2007). Specific correlations have emerged in response to temperature, solar radiation, and rainfall. One study shows that increased temperature leads to increased concentrations of all proteinogenic amino acids (Carrera et al., 2011), while another concludes that only methionine and cysteine increase alongside temperature (Wolf et al., 1982). Increased solar radiation and greater available water appeared to have a negative relationship with amino acid content (Carrera et al., 2011). These favorable conditions would increase yield which has been shown to have a negative correlation with overall protein content (Wilcox and Shibles, 2001). The multitude of agronomic factors that affect amino acid profiles in soybean make it exceedingly important to compensate for variables when researching.

Processing Impacts

The diversity in food, feed, and industrial used for soybean require the whole seed or seed components to be processed. Processing can affect the nutritional value of soybean protein and presence of amino acids in food and feed. Processing procedures can either separate seed components for different purposes or convert the entire seed into a product (usually human food). Some human soy foods such as edamame and soybean sprouts need little to no processing. Others including soymilk, tofu, natto, and soy sauce involve more processing. Soymilk and tofu processing are interconnected. Soymilk is a water-extract of whole or crushed soybeans that is coagulated and pressed into tofu (Liu,

1997). While not all seed proteins convert into protein in tofu, 11S/7S storage protein ratios have been shown to be both positively and negatively correlated with tofu hardness (Cai and Chang, 1999; Mujoo et al., 2003). Natto is a soy food created by fermenting whole soybeans with Bacillus subtilis. Fermentation time affects final amino acid concentrations, and proper fermentation length could potentially increase nutritional values (Weng and Chen, 2010). Soy sauce is produced by traditional and commercial methods, but both are based around whole seed or meal fermentation with Aspergillus sp. However, commercial methods have a lower amino acid to nitrogen ratio (Liu, 1997).

Soybean meal processing also impacts the level of amino acids in livestock feed. The first step in soybean meal processing is essentially separating protein from oil. A variety of methods exist including solvent extraction, screw pressing, and extruding (Liu, 1997; Lusas, 2004; Johnson and Smith,). All three processes have three final products: oil, meal (usually toasted to lessen anti-nutritional factors), and hulls. Over processing of solvent extracted soymeal has been shown to decrease lysine, cysteine, and arginine levels (Taira, 1966; Parsons et al., 1992). Protein solubility and dispersibility measurements may be a useful indicator of over processing (Araba and Dale, 1990; Batal et al., 2000). Soybean hulls are sometimes added to livestock feed for additional fiber, however an increase in hull/meal ratios decrease the digestibility of amino acids (Dilger et al., 2004). While soybean is renowned for its protein and amino acid content, actual nutritional values can be decreased through certain processing methods.

Evaluation Methods

All previously mentioned aspects of soybean production in regard to amino acid levels and human and animal nutrition depend on a single common denominator: amino

acid quantification. Amino acids must be reliably, effectively, and accurately identified, measured and evaluated. A Google Scholar search of "amino acid analysis" will display over 1 million results. Several reviews have been published regarding the development of amino acid analysis (Tristram and Rattenbury, 1981; Williams, 1994; Husek and Simek, 2001). In general, contemporary analysis of amino acids from any source will be performed by chromatography or near-infrared reflectance spectroscopy. Chromatography is the common method with specific techniques including ion exchange chromatography (IEC), high-performance liquid chromatography (HPLC), and gas chromatography (GC). HPLC is the more validated method for soybean amino acid analysis. It is more efficient than IEC, and it does not require the transformation into volatiles like GC (Malmer and Schroeder, 1990; Oomah et al., 1994; Williams, 1994; Jajić et al., 2013). Near-infrared reflectance spectroscopy (NIRS) is a more recent addition to amino acid analysis, and it has the potential to drastically improve the efficiency in soybean feed evaluation [United Soybean Board(a)]. The inability to actually measure amino acid levels is main hindrance for NIRS amino acid analysis. NIRS methods must be developed from a calibration set of raw data (often from HPLC) (Kovalenko et al., 2006; Baianu and Prisecaru, 2011; Pazdernik et al.,). Nonetheless, efficiency improvements should persuade researchers to continually explore future NIRS amino acid analysis applications.

Conclusion

Soybean is a valuable source of protein and amino acids for humans and livestock. Soybean's well-balanced amino acid profile provides all essential amino acids as well as most non-essential. However, there is much room for nutritional improvement.

Proteinogenic sulfur-containing amino acids, methionine and cysteine, are deficient in soybean and are especially needed in livestock rations. Increased levels of these amino acids would augment soybean meal value and lessen the need for synthetic amino acid supplements. Breeding efforts have made little progress in adjusting amino acid profiles thus far, however significant developments in understanding genomic control regions promise future success. Genetic engineering efforts have shown promising amino acid improvements, but regulations and public opinions made commercialization difficult. New gene-editing technology could be the key to unlock true nutritional improvement.

Agronomic properties and processing methods both impact the final quantities of amino acids available to humans and livestock. Understanding these impacts are essential to improve the nutritional quality of soybeans. Amino acid evaluation through HPLC provides reliable and efficient quantification, yet even quicker measurements are possible through NIRS. As the world's population continues to grow, soybeans will be essential to both human and livestock for amino acid requirements. Wholesome approaches that understand the complexities of amino acids in soybean will be required to maximize overall success and feed the world with balance soy proteins.

References

Allee, M.S.K. and G.L. 2003. Modifications in Soybean Seed Composition to Enhance Animal Feed Use and Value: Moving From a Dietary Ingredient to a Functional Dietary Component. http://agbioforum.org/v6n12/v6n12a05-kerley.htm (accessed 4 August 2019).

Araba, M., and N.M. Dale. 1990. Evaluation of Protein Solubility as an Indicator of Overprocessing Soybean Meal. Poult. Sci. 69(1): 76–83. doi: 10.3382/ps.0690076.

Arora, S.K., and Y.P. Luthra. 1971. Relationship between sulfur content of leaf with methionine, cysteine and cysteine contents in the seeds ofPhaseolus aureus L. As affected by S, P, and N application. Plant Soil 34(1): 91–96. doi: 10.1007/BF01372763.

Baianu, I., and V. Prisecaru. 2011. Novel NIR Spectroscopy Correlation Approach to Amino Acid Analysis of Soybean Proteins for Composition Improvements. Nat. Preced. doi: 10.1038/npre.2011.6231.1.

Batal, A.B., M.W. Douglas, A.E. Engram, and C.M. Parsons. 2000. Protein dispersibility index as an indicator of adequately processed soybean meal. Poult. Sci. 79(11): 1592–1596. doi: 10.1093/ps/79.11.1592.

Berry, T.H., D.E. Becker, O.G. Rasmussen, A.H. Jensen, and H.W. Norton. 1962. The Limiting Amino Acids in Soybean Protein. J. Anim. Sci. 21(3): 558–561. doi: 10.2527/jas1962.213558x.

Boisen, S., T. Hvelplund, and M.R. Weisbjerg. 2000. Ideal amino acid profiles as a basis for feed protein evaluation. Livest. Prod. Sci. 64(2): 239–251. doi: 10.1016/S0301-6226(99)00146-3.

Bonato, M.A., N.K. Sakomura, J.C. Siqueira, J.B.K. Fernandes, and R.M. Gous. 2011. Maintenance requirements for methionine and cysteine, and threonine for poultry. South Afr. J. Anim. Sci. 41(3): 209-222–222. doi: 10.4314/sajas.v41i3.3.

Brosnan, J.T., and M.E. Brosnan. 2006. The Sulfur-Containing Amino Acids: An Overview. J. Nutr. 136(6): 1636S-1640S. doi: 10.1093/jn/136.6.1636S.

Bunchasak, C. 2009. Role of Dietary Methionine in Poultry Production. J. Poult. Sci. 46(3): 169–179. doi: 10.2141/jpsa.46.169.

Buttery, P.J., and J.P.F. D'Mello. 1994. Amino Acid Metabolism in Farm Animals: An Overview. Amino Acids in Farm Animal Nutrition. CAB International, Wallingford, United Kingdom. p. 1–10

Cai, T., and K.-C. Chang. 1999. Processing Effect on Soybean Storage Proteins and Their Relationship with Tofu Quality. J. Agric. Food Chem. 47(2): 720–727. doi: 10.1021/jf980571z.

Carrera, C.S., C.M. Reynoso, G.J. Funes, M.J. Martínez, J. Dardanelli, et al. 2011. Amino Acid Composition of Soybean Seeds as Affected by Climatic Variables. Braz. Agric. Res. 46(12): 1579–1587. doi: 10.1590/S0100-204X2011001200001.

Catoni, G.L. 1953. S-Adenosylmethionine; a new intermediate formed enzymatically from L-methionine and adenosinetriphosphate. J. Biol. Chem. 204(1): 403–416.

Conde-Aguilera, J.A., C. Cobo-Ortega, S. Tesseraud, M. Lessire, Y. Mercier, et al. 2013. Changes in body composition in broilers by a sulfur amino acid deficiency during growth. Poult. Sci. 92(5): 1266–1275. doi: 10.3382/ps.2012-02796.

Dilger, R.N., J.S. Sands, D. Ragland, and O. Adeola. 2004. Digestibility of nitrogen and amino acids in soybean meal with added soyhulls,. J. Anim. Sci. 82(3): 715–724. doi: 10.2527/2004.823715x.

D'Mello, J.P.F. 1994. Responses of Growing Poultry to Amino Acids. Amino Acids in Farm Animal Nutrition. CAB International, Wallingford, United Kingdom. p. 205–244

D'Mello, J.P.F., editor. 2012. Amino Acids in Human Nutrition and Health. CAB International, Wallingford, United Kingdom.

Drovers. Control Feed Costs with Amino Acids. https://www.drovers.com/article/control-feed-costs-amino-acids (accessed 5 August 2019).

El-Shemy, H., M. Khalafalla, K. Fujita, and M. Ishimoto. 2007. Improvement of protein quality in transgenic soybean plants. Biol. Plant. 51: 277–284. doi: 10.1007/s10535-007-0055-3.

Falco, S.C., T. Guida, M. Locke, J. Mauvais, C. Sanders, et al. 1995. Transgenic Canola and Soybean Seeds with Increased Lysine. Bio/Technology 13(6): 577. doi: 10.1038/nbt0695-577.

Fallen, B., C. Hatcher, F. Allen, D. Kopsell, A. Saxton, et al. 2013. Soybean Seed Amino Acid Content QTL Detected Using the Universal Soy Linkage Panel 1.0 with 1,536 SNPs. Agron. Hortic. -- Fac. Publ. https://digitalcommons.unl.edu/agronomyfacpub/807.

Fernandez, S.R., S. Aoyagi, Y. Han, C.M. Parsons, and D.H. Baker. 1994. Limiting order of amino acids in corn and soybean meal for growth of the chick. Poult. Sci. 73(12): 1887–1896. doi: 10.3382/ps.0731887.

Fontecave, M., M. Atta, and E. Mulliez. 2004. S-adenosylmethionine: nothing goes to waste. Trends Biochem. Sci. 29(5): 243–249. doi: 10.1016/j.tibs.2004.03.007.

Fuller, M.F. 1994. Amino Acid Requirements for Maintenance, Body Protein Accretion and Reproduction in Pigs. Amino Acids in Farm Animal Nutrition. CAB International, Wallingford, United Kingdom. p. 155–184

Gayler, K.R., and G.E. Sykes. 1985. Effects of Nutritional Stress on the Storage Proteins of Soybeans. Plant Physiol. 78(3): 582–585. doi: 10.1104/pp.78.3.582.

Gioseffi, E., A. de Neergaard, and J.K. Schjoerring. 2012. Interactions between uptake of amino acids and inorganic nitrogen in wheat plants. Biogeosciences 9(4): 1509–1518. doi: https://doi.org/10.5194/bg-9-1509-2012.

Goldflus, F., M. Ceccantini, and W. Santos. 2006. Amino acid content of soybean samples collected in different Brazilian states. Braz. J. Poult. Sci. 8(2): 105–111. doi: 10.1590/S1516-635X2006000200006.

Haghiri, F. 1966. Influence of Macronutrient Elements on the Amino Acid Composition of Soybean Plants. Agron. J. 58.

Han, Y., H. Suzuki, C.M. Parsons, and D.H. Baker. 1992. Amino Acid Fortification of a Low-Protein Corn and Soybean Meal Diet for Chicks. Poult. Sci. 71(7): 1168–1178. doi: 10.3382/ps.0711168.

Hermann, F.J. 1962. A revision of the genus Glycine and its immediate allies. U.S. Dept. of Agriculture.

Herrmann, K. 1983. The Common Aromatic Biosynthetic Pathway. Amino Acids: Biosynthesis and Genetic Regulation. Addison-Wesley Publishing Company, Inc., Reading, Massachusetts

Herrmann, K., and R.L. Somerville, editors. 1983. Amino Acids: Biosynthesis and Genetic Regulation. Addison-Wesley Publishing Company, Inc., Reading, Massachusetts.

Howell, R.W., C.A. Brim, and R.W. Rinne. 1972. The plant geneticist's contribution toward changing lipid and amino acid composition of soybeans. J. Am. Oil Chem. Soc. 49(1): 30–32. doi: 10.1007/BF02545134.

Husek, P., and P. Simek. 2001. Advances in Amino Acid Analysis. LCGC 19(9): 986–999.

Ingenbleek, Y., and H. Kimura. 2013. Nutritional essentiality of sulfur in health and disease. Nutr. Rev. 71(7): 413–432. doi: 10.1111/nure.12050.

Jajić, I., S. Krstović, D. Glamočić, S.M. Jakšić, and B.F. Abramović. 2013. Validation of an HPLC method for the determination of amino acids in feed. J. Serbian Chem. Soc. 78(6): 839–850. doi: 10.2298/jsc120712144j.

Jankowski, J., M. Kubińska, and Z. Zduńczyk. 2014. Nutritional and immunomodulatory function of methionine in poultry diets – a review. Ann. Anim. Sci. 14(1): 17–32. doi: 10.2478/aoas-2013-0081.

Jessop, C.E., S. Chakravarthi, R.H. Watkins, and N.J. Bulleid. 2004. Oxidative protein folding in the mammalian endoplasmic reticulum. Biochem. Soc. Trans. 32(Pt 5): 655–658. doi: 10.1042/BST0320655.

Jez, J.M., and N.K. Fukagawa. 2008. Plant sulfur Compounds and Human Health. Sulfur: A Missing Link Between Soils, Crops, and Nutrition. p. 281–291

Johnson, L., and K. Smith. Soybean Processing. United Soybean Board. https://www.soymeal.org/wp-content/uploads/2018/04/soybean_processing.pdf (accessed 11 August 2019).

Kapoora, A.C., and Y.P. Gupta. 1977. Changes in proteins and amino acids in developing soybean seed and effect of phosphorus nutrition. J. Sci. Food Agric. 28(2): 113–120. doi: 10.1002/jsfa.2740280202.

Kerr, P. 1996. Utilization and Quality of Identity Preserved Oilseed Co-products. New Orleans, LA

Kim, W.-S., and H.B. Krishnan. 2019. Impact of co-expression of maize 11 and 18 kDa δ-zeins and 27 kDa γ-zein in transgenic soybeans on protein body structure and sulfur amino acid content. Plant Sci. 280: 340–347. doi: 10.1016/j.plantsci.2018.12.016.

Kita, Y., Y. Nakamoto, M. Takahashi, K. Kitamura, K. Wakasa, et al. 2010. Manipulation of amino acid composition in soybean seeds by the combination of deregulated tryptophan biosynthesis and storage protein deficiency. Plant Cell Rep. 29(1): 87–95. doi: 10.1007/s00299-009-0800-5.

Kovalenko, I.V., G.R. Rippke, and C.R. Hurburgh. 2006. Determination of Amino Acid Composition of Soybeans (Glycine max) by Near-Infrared Spectroscopy. J. Agric. Food Chem. 54(10): 3485–3491. doi: 10.1021/jf052570u.

Kredich, N.M. 1983. Regulation of Cysteine Biosynthesis in Escherichia coli and Salmonella typhimurium. Amino Acids: Biosynthesis and Genetic Regulation. Addison-Wesley Publishing Company, Inc., Reading, Massachusetts. p. 115–132

Kuiken, K.A., and M. Lyman. 1949. Essential amino acid composition of soybean meals prepared from twenty strains of soybeans. J Biol Chem (177): 29–36.

Levine, R.L., L. Mosoni, B.S. Berlett, and E.R. Stadtman. 1996. Methionine residues as endogenous antioxidants in proteins. Proc. Natl. Acad. Sci. U. S. A. 93(26): 15036–15040. doi: 10.1073/pnas.93.26.15036.

Lewis, A.J. 2003. Methionine-cysteine Relationships in Pig Nutrition. Absorption and Utilization of Amino Acids. CAB International, Wallingford, United Kingdom. p. 143–156

Li, X., R. Tian, S. Kamala, H. Du, W. Li, et al. 2018. Identification and verification of pleiotropic QTL controlling multiple amino acid contents in soybean seed. Euphytica 214(6): 93. doi: 10.1007/s10681-018-2170-y.

Liu, K. 1997. Soybeans: Chemistry, Technology, and Utilization. Chapman & Hall, New York, NY.

Lusas, E.W. 2004. Soybean Processing and Utilization. Soybeans: Improvement, Production, and Uses. 3rd ed. ASA, CSSA, SSSA, Madison, WI. p. 949–1036

Mahmoud, A.A., S.S. Natarajan, J.O. Bennett, T.P. Mawhinney, W.J. Wiebold, et al. 2006. Effect of Six Decades of Selective Breeding on Soybean Protein Composition and Quality: A Biochemical and Molecular Analysis. J. Agric. Food Chem. 54(11): 3916–3922. doi: 10.1021/jf060391m.

Malmer, M.F., and L.A. Schroeder. 1990. Amino acid anslysis by high-performance liquid chromatography with methanesulfonic acid hydrolysis and 9-fluorenylmethylchloroformate derivatization. J. Chromatogr. A 514: 227–239. doi: 10.1016/S0021-9673(01)89394-8.

Mato, J.M., M.L. Martinez-Chantar, and S.C. Lu. 2012. Methionine Metabolism. Amino Acids in Human Nutrition and Health. CAB International, Wallingford, United Kingdom. p. 173–188

Meinke, D.W., J. Chen, and R.N. Beachy. 1981. Expression of storage-protein genes during soybean seed development | SpringerLink. Planta (153): 130–139.

Meisinger, J.J. 2005. 1 Ammonia Volatilization from Dairy and Poultry Manure

Moskovitz, J. 2005. Methionine sulfoxide reductases: ubiquitous enzymes involved in antioxidant defense, protein regulation, and prevention of aging-associated diseases. Biochim. Biophys. Acta 1703(2): 213–219. doi: 10.1016/j.bbapap.2004.09.003.

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

Oomah, B.D., H. Voldeng, and J.A. Fregeau-Reid. 1994. Characterization of soybean proteins by HPLC. Plant Foods Hum. Nutr. (45): 251–263.

Osborne, T.B., and L.B. Mendel. 1993. Amino-acids in nutrition and growth. 1914. J. Am. Coll. Nutr. 12(5): 484–485. doi: 10.1080/07315724.1993.10718340.

Paek, N.C., J. Imsande, R.C. Shoemaker, and R. Shibles. 1997. Nutritional Control of Soybean Seed Storage Protein. Crop Sci. 37(2): 498–503. doi: 10.2135/cropsci1997.0011183X003700020031x.

Panthee, D., and V.R. Pantalone. 2006. Registration of Soybean Germplasm Lines TN03–350 and TN04–5321 with Improved Protein Concentration and Quality. Crop Sci. - CROP SCI 46. doi: 10.2135/cropsci2005.11.0437.

Panthee, D.R., V.R. Pantalone, C.E. Sams, A.M. Saxton, D.R. West, et al. 2006a. Quantitative trait loci controlling sulfur containing amino acids, methionine and cysteine, in soybean seeds. Theor. Appl. Genet. 112(3): 546–553. doi: 10.1007/s00122-005-0161-6.

Panthee, D.R., V.R. Pantalone, A.M. Saxton, D.R. West, and C.E. Sams. 2006b. Genomic regions associated with amino acid composition in soybean. Mol. Breed.: 79–89. doi: 10.1007/s11032-005-2519-5.

Parsons, C.M., K. Hashimoto, K.J. Wedekind, Y. Han, and D.H. Baker. 1992. Effect of Overprocessing on Availability of Amino Acids and Energy in Soybean Meal. Poult. Sci. 71(1): 133–140. doi: 10.3382/ps.0710133.

Pazdernik, D.L., A.S. Killam, and J.H. Orf. Analysis of Amino and Fatty Acid Composition in Soybean Seed, Using Near Infrared Refelctance Spectroscopy. Agron. J. 89: 679–685.

Pfarr, M.D., M.J. Kazula, J.E. Miller-Garvin, and S.L. Naeve. 2018. Amino Acid Balance is Affected by Protein Concentration in Soybean. Crop Sci. 58(5): 2050–2062. doi: 10.2135/cropsci2017.11.0703.

Rackis, J.J., R.L. Anderson, H.A. Sasame, A.K. Smith, and C.H. VanEtten. 1961. Soybean Amino Acids, Amino Acids in Soybean Hulls and Oil Meal Fractions. J. Agric. Food Chem. 9(5): 409–412. doi: 10.1021/jf60117a021.

Schmidt, M.A., W.B. Barbazuk, M. Sandford, G. May, Z. Song, et al. 2011. Silencing of Soybean Seed Storage Proteins Results in a Rebalanced Protein Composition Preserving Seed Protein Content without Major Collateral Changes in the Metabolome and Transcriptome. Plant Physiol. 156(1): 330–345. doi: 10.1104/pp.111.173807.

Sleper, D.A., and J.M. Poehlman. 2006. Breeding Field Crops. 5th ed. Blackwell Publishing Professional, Ames, Iowa.

Taira, H. 1966. Studies on Amino Acid Contents of Processed Soybean: Part X. The Influence of Added Sugars on the Heat Destruction of the Basic and Sulfur Containing Amino Acids in Soybean Products. Agric. Biol. Chem. 30(9): 847–855. doi: 10.1080/00021369.1966.10858695.

Takahashi, M., Y. Uematsu, K. Kashiwaba, K. Yagasaki, M. Hajika, et al. 2003. Accumulation of high levels of free amino acids in soybean seeds through integration of mutations

conferring seed protein deficiency. Planta 217(4): 577–586. doi: 10.1007/s00425-003-1026-3.

Teixeira, W.F., E.B. Fagan, L.H. Soares, R.C. Umburanas, K. Reichardt, et al. 2017. Foliar and Seed Application of Amino Acids Affects the Antioxidant Metabolism of the Soybean Crop. Front. Plant Sci. 8. doi: 10.3389/fpls.2017.00327.

Thakur, M., and C.R. Hurburgh. 2007. Quality of US Soybean Meal Compared to the Quality of Soybean Meal from Other Origins. J. Am. Oil Chem. Soc. 84(9): 835–843. doi: 10.1007/s11746-007-1107-8.

The American Soybean Association. 2019a. International: World Oilseed Production. http://soystats.com/international-world-oilseed-production/ (accessed 10 July 2019).

The American Soybean Association. 2019b. International: World Soybean Production. http://soystats.com/international-world-soybean-production/ (accessed 17 July 2019).

The American Soybean Association. 2019c. U.S. Exports: Soy Exports by Customer. http://soystats.com/u-s-exports-soy-exports-by-customersda/ (accessed 10 July 2019).

Townsend, J.A., L.A. Thomas, E.S. Kulisek, M.J. Daywalt, K.R.K. Winter, et al. 1992. Improving the Quality of of Seed Proteins in Soybean. Proceedings of the 4th Biennial Conference on Molecular and Cellular Biology of Soybean. Iowa State University, Ames, Iowa. p. 4

Tristram, G.R., and J.M. Rattenbury. 1981. The Development of Amino Acid Analysis. Amino Acid Analysis. Ellis Horwood, Chichester, UK. p. 16–36

United Soybean Board(a). Measuring Beyond the Bushel. https://unitedsoybean.org/article/measuring-beyond-the-bushel-1/ (accessed 10 July 2019).

United Soybean Board(b). Soybean Meal. https://unitedsoybean.org/topics/soybean-meal/ (accessed 17 July 2019).

Wade, C. 1985. Carlson Wade's Amino Acids Book. Keats Publishing Inc., New Canaan, Connecticut.

Wagner, I., and H. Musso. 1983. New Naturally Occurring Amino Acids. Angew. Chem. Int. Ed. Engl. 22(11): 816–828. doi: 10.1002/anie.198308161.

Waldroup, P.W., R.J. Mitchell, J.R. Payne, and K.R. Hazen. 1976. Performance of chicks fed diets formulated to minimize excess levels of essential amino acids. Poult. Sci. 55(1): 243–253. doi: 10.3382/ps.0550243.

Warrington, C.V., H. Abdel-Haleem, D.L. Hyten, P.B. Cregan, J.H. Orf, et al. 2015. QTL for seed protein and amino acids in the Benning × Danbaekkong soybean population. Theor. Appl. Genet. 128(5): 839–850. doi: 10.1007/s00122-015-2474-4.

Weaver, R. 2008. Molecular Biology. 4th ed. McGraw-Hill, New York, NY.

Weng, T.M., and M.T. Chen. 2010. Changes of Protein in Natto (a fermented soybean food) Affected by Fermenting Time. Food Sci. Technol. Res. 16(6): 537–542. doi: 10.3136/fstr.16.537.

Wilcox, J.R., and R.M. Shibles. 2001. Interrelationships among Seed Quality Attributes in Soybean. Crop Sci. 41(1): 11–14. doi: 10.2135/cropsci2001.41111x.

Williams, A.P. 1994. Recent Developments in Amino Acid Analysis. Amino Acids in Farm Animal Nutrition. CAB International, Wallingford, United Kingdom. p. 11–36

Willke, T. 2014. Methionine production—a critical review. Appl. Microbiol. Biotechnol. 98(24): 9893–9914. doi: 10.1007/s00253-014-6156-y.

Wolf, W.J. 1969. Soybean protein nomenclature: a progress report. https://pubag.nal.usda.gov/catalog/31396 (accessed 9 August 2019).

Wolf, R.B., J.F. Cavins, R. Kleiman, and L.T. Black. 1982. Effect of temperature on soybean seed constituents: Oil, protein, moisture, fatty acids, amino acids and sugars. J. Am. Oil Chem. Soc. 59(5): 230–232. doi: 10.1007/BF02582182.

Wu, G. 2014. Dietary requirements of synthesizable amino acids by animals: a paradigm shift in protein nutrition. J. Anim. Sci. Biotechnol. 5(1): 34. doi: 10.1186/2049-1891-5-34.

Wu, B., H. Cui, X. Peng, J. Fang, W. Cui, et al. 2012. Effect of Methionine Deficiency on the Thymus and the Subsets and Proliferation of Peripheral Blood T-Cell, and Serum IL-2 Contents in Broilers. J. Integr. Agric. 11(6): 1009–1019. doi: 10.1016/S2095-3119(12)60093-8.

Wylie, L.M., G.W. Robertson, M.G. Macleod, and P.M. Hocking. 2001. Effects of ambient temperature and restricted feeding on the growth of feathers in growing turkeys. Br. Poult. Sci. 42(4): 449–455. doi: 10.1080/00071660120070631.

Chapter 2: Genome-wide Association Study and Genomic Selection for Proteinogenic Methionine in Soybean Seeds

William M. Singer[1], Zachary Shea[1], Dajun Yu[2], Haibo Huang[2], Rouf Mian[3], Chao Shang[1], Maria L. Rosso[1], Qijian Song[4], and Bo Zhang[1]

[1]School of Plant and Environmental Sciences, Virginia Tech

[2]Department of Food Science and Technology, Virginia Tech

[3]Soybean & Nitrogen Fixation Unit, USDA-ARS

[4]Soybean Genomics and Improvement Laboratory, Beltsville Agricultural Research Center, USDA-ARS

Abstract

Soybean [*Glycine max (L.) Merr.*] seeds have an amino acid profile that provides excellent viability as a food and feed protein source. However, low concentrations of an essential amino acid, methionine, limit the nutritional utility of soybean protein. The objectives of this study were to identify genomic associations and evaluate the potential for genomic selection (GS) for methionine content in soybean seeds. We performed a genome-wide association study (GWAS) that utilized 311 soybean accession from maturity groups IV and V grown in three locations in 2018 and 2019. A total of 35,570 single nucleotide polymorphisms (SNPs) were used to identify genomic associations with proteinogenic methionine content that was quantified by high-performance liquid chromatography (HPLC). Across four environments, 23 novel SNPs were identified as being associated with methionine content. The strongest associations were found on chromosomes 3, 8, and 16, and several gene models were recognized within proximity to these SNPs, such as a leucine-rich repeat protein kinase and a serine/threonine protein kinase. Identification of these SNPs should help elucidate genomic regions for use by soybean breeders to improve protein quality in soybean seed. GS was evaluated using k-fold cross validation within each environment with two SNP sets, the complete 35,570 set and a subset of 248 SNPs determined to be associated with methionine through GWAS. Average prediction accuracy (r^2) was highest using the SNP subset ranging from 0.45-0.62, which was a significant improvement from the complete set accuracy that ranged from 0.03-0.27. This suggests GS that utilizes a significant subset of SNPs may be a viable tool for soybean breeders seeking to improve methionine content.

Introduction

Soybean [*Glycine max (L.) Merr.*] has an ideal amino acid profile for a protein source used in livestock feed and human food. All nine essential amino acids, histidine (His), isoleucine (Ile) leucine (Leu), lysine (Lys), methionine (Met), phenylalanine (Phe), threonine (Thr), tryptophan (Trp), and valine (Val), are present in soybean seeds (Kuiken and Lyman, 1949; Boisen et al., 2000). Accounting for 35% of the seed (Wilson, 2004), the protein component is processed into meal and regularly used in cattle, swine, and poultry feed (Buttery and D'Mello, 1994). During 2020, 33.2 million metric tons of soybean meal were used in the United States for livestock feed, in which 20.2, 6.3, and 5.8 million metric tons were fed to poultry, swine, and cattle, respectively (The American Soybean Association, 2020).

While all essential amino acids are present, soybean is deficient in Met which limits its nutritional utility in feed (Berry et al., 1962; Fernandez et al., 1994; Bonato et al., 2011). Met is required for metabolic processes and is the initiating amino acid in protein synthesis (Brosnan et al., 2007). Due to Met deficiency, poultry has displayed negative effects on body composition such as protein, fat, and tissue gain (Conde-Aguilera et al., 2013) and disease immunity (Wu, 2014). For this reason, synthetic supplementation of Met is critical to livestock feed, especially poultry. Bunchasak (2009) summarized the importance, viability, and special considerations for Met supplementation, however, synthetic methionine production generates hazardous waste and contributes to the greater dependence on fossil fuels (Willke, 2014; Neubauer and Landecker, 2021). Therefore, a sustainable solution would be to increase Met concentrations in soybean protein content through plant breeding.

Since soybean was introduced to North America in 1765 (Hymowitz and Harlan, 1983), it has gained global prevalence. Contemporary soybean breeders have dedicated enormous effort to improve seed composition. Patil et al. (2017) aptly reviewed and described modern genomic efforts to improve soybean protein content. More specifically, quantitative trait loci (QTL) have been identified for protein concentration (Panthee et al., 2005; Warrington et al., 2015) as well as amino acid profiles (Panthee et al., 2006a; Panthee et al., 2006b; Fallen et al., 2013; Warrington et al., 2015; Li et al., 2018). Direct breeding results from this research include the sole publicly-developed United States soybean variety (TN04–5321) release with enhanced sulfur-containing amino acids concentrations (Panthee and Pantalone, 2006) and potential introgression of an allele for significantly increased protein content (Warrington et al., 2015). Additionally, recent advances in molecular markers and high-throughput sequencing, summarized well by Zargar et al. (2015), have allowed for genomic research at the genome-wide level. Hwang et al. (2014) and Li et al. (2019) used single nucleotide polymorphisms (SNPs) to pinpoint genetic control of protein in soybean seed through genome-wide association studies (GWAS). Lee et al. (2019) targeted protein content as well as four amino acids, Met, Cys, Lys, and Thr, through GWAS. Qin et al. (2019) used GWAS to find genomic associations for 15 amino acids, Ala, Arg, Asp, Glu, Gly, His, Ile, Leu, Lys, Phe, Pro, Ser, Thr, Tyr, and Val. A single study also focused directly on Met and Cys with genome-wide associations for Canadian soybean lines in MG 000-II (Malle et al., 2020). Lee et al. (2019) and Malle et al. (2020) reported Met measurements using near-infrared reflectance spectroscopy (NIRS), whereas Qin et al. (2019) utilized ion exchange chromatography.

Genomic selection (GS) utilizes similar statistical models as GWAS, but it seeks to exploit larger genomic variations than individual genomic regions (Meuwissen et al., 2001). GS has been shown to reduce selection time in soybean breeding (Matei et al., 2018) and the U.S. soybean germplasm collection has proven to be a valuable resource for creating GS models (Jarquin et al., 2016). Promising results have displayed successful prediction of grain yield, protein and oil content, plant height, maturity, seed weight (Ma et al., 2016; Duhnen et al., 2017; Stewart-Brown et al., 2019; Ravelombola et al., 2021) as well as soybean cyst nematode resistance (Ravelombola et al., 2019, 2020). However, only one study by Qin et al. (2019) has evaluated GS for amino acid content in soybean seed, and it did not include Met concentrations.

Additionally, Warrington et al. (2015) identified negative correlations between increased protein content and Lys, Thr, and Met+Cys concentrations. This suggests complex genetic controls of protein as soybean breeders balance objectives for protein quantity and quality moving forward. Therefore, this project seeks to further elucidate genomic associations through GWAS and evaluate the potential for GS of proteinogenic Met content in soybean seeds.

Materials and Methods

Plant Materials

A total of 500 soybean accessions were selected from the USDA Soybean Germplasm Collection to represent maximum genetic variability in maturity groups IV and V based on genetic distance (Qin et al., 2017). Among them, a panel consisting of 311 accessions from 17 different countries (Table 1) with good seed quality were grown in 3 m two-row plots with 76 cm row spacing in Blacksburg, VA and 4.2 m single row plots with 96 cm row spacing in Clayton, NC in 2018. In 2019, they were grown in 3 m four-row plots with 76 cm row spacing in Warsaw, VA and repeated in Blacksburg, VA. Plots were organized based upon maturity and grown as a randomized complete block design (RCBD) with two blocks at each location. Each block included two commercial checks AG4403 and Ellis. Due to limited seed quantity in general, block replicates were merged prior to seed processing.

Data Collection

All seed samples were cleaned by removing moldy, mottled, discolored, or off-types seeds. Dry-matter based protein content and moisture were measured using the DA 7250 NIR Analyzer spectrophotometer (PerkinElmer Inc.) through near-infrared reflectance spectroscopy (NIRS). For NIRS, the manufacturer's annual updated calibration module was used and protein content was recorded for each sample.

Samples were ground using a water-cooler Foss 1095 Knifetec mill to a consistent particle size. Subsamples of 0.01g were weighed into glass digestion tubes and subsequently hydrolyzed using a modified method 994.12 (AOAC International) to break apart proteinogenic methionine. Samples were first oxidized with 0.5 mL of performic

acid at 0˚C for 16 hours and 200 µL of sodium metabisulfite solution was added to end the reaction. Hydrolysis was then performed with 3 mL of 6 M HCl at 110˚C for 16 hours. Next, samples were diluted to 10 mL with water, and 750 µL subsamples were taken and concentrated to remove HCl.

Concentrated samples were rehydrated with water into vials for high-performance liquid chromatography (HPLC). HPLC was performed using online derivatization with o-phthalaldehyde (OPA), ultra-violet (UV) detection, and the Agilent AdvanceBio AAA column with Agilent HPLC model 1200. Each sample had two technical replicates that were averaged to account for biological and equipment variation. To better describe proteinogenic concentrations, Met was reported on a g/kg crude protein (g kg^{-1} cp) basis. Data were fit with an ANOVA using standard least squares that included accession, location, and year as fixed effects.

Genotypic Data

Publicly available SNP marker data (www.soybase.org) of the 311 accessions were downloaded from the SoySNP50K SNPs data repository (Song et al., 2015). A total of 42,509 initial SNPs were filtered by low minor allele frequency (MAF < 0.05) and missing genotypes, which resulted in 35,570 SNPs being used for further analysis.

Population Structure

Population structure was evaluated through a discriminant analysis of principal components (DAPC) using the adegenet package (Jombart, 2008) in R to identify clusters of genetically related individuals (Jombart et al., 2010). Successive k-means clustering with the function find.clusters with maximum clusters as k = 40 was used. A total of 300 principal components were retained, and Bayesian information criterion (BIC) was used

to identify an optimal number of clusters. The function dapc was then used by retaining an optimal number of principal components to maximize cumulative variance without overfitting, and all discriminant functions and eigenvalues were retained. A kinship matrix was also created with the software TASSEL 5 (Bradbury et al., 2007) using the Centered_IBS method (Endelman and Jannink, 2012).

Genome-wide Association Analysis and Candidate Gene Evaluation

Associations between genotypic and phenotypic data were analyzed using two different models in TASSEL 5: mixed linear model (MLM) and general linear model (GLM). Predominantly, MLM was used to incorporate a kinship matrix (K) jointly with population structure (Q) for increased statistical power through the Q+K approach (Yu et al., 2006). GLM was used to examine individual location datasets through a more lenient least squares fixed effect model with Q as a covariate. Additionally, five principal components (accounting for 18.75% cumulative variance) were included as covariates for the 2018 Blacksburg, VA and 2019 Warsaw, VA datasets to better control for false positive associations. A modified Šidák correction $\alpha_{sid} = 1 - (1 - \alpha)^{1/m}$ for multiple testing was used to identify significant associations. The effective number of markers (Meff) was calculated to be 4,191 using the poolr package in R with the Li and Ji method (Li and Ji, 2005). Meff replaced m, and thus, the adjusted significance threshold at α = 5% and the suggestive threshold at α = 25% were $-\log_{10}(P) > 4.91$ and $-\log_{10}(P) > 4.16$, respectively. QQ and Manhattan plots were used to visualize results with the qqman package (Turner, 2014). Gene models from Glyma.Wm82.a2.v1 (Williams 82) as displayed on www.soybase.com within 10 kb of significant SNPs flanking regions were reported as candidate genes (Xie et al., 2018; Qin et al., 2019).

Gene descriptions were reported from gene homolog descriptions from TAIR for Arabidopsis thaliana (Berardini et al., 2015). If TAIR homologs were not available, descriptions were reported from either PANTHER or GO databases (Ashburner et al., 2000; Mi et al., 2013; Gene Ontology Consortium, 2021). Expression patterns within soybean reproductive tissues (flowers, pods, and seeds) of each gene model were also reported when available (Severin et al., 2010).

Genomic Selection

GS was performed using gBLUP (genomic best linear unbiased prediction) with the TASSEL 5 genomic selection function. Similar to the GWAS, the Q+K approach was used to fit a mixed model with population structure and a kinship matrix as covariates. K-fold cross validation was performed using k = 5 with 20 iterations, and the Pearson's correlation coefficient (r^2) was collected for each fold. Each environment's dataset underwent GS using all 35,570 SNPs as well as a subset of 248 SNPs identified as having $-\log_{10}(P) > 3$ from the GWAS (Qin et al., 2019). A T-test was used to compare r^2 values between the whole and partial SNP models.

Results

Phenotypic

Methionine concentrations across all environments displayed normal, continuous distributions with a grand mean of 9.06 g kg $^{-1}$ cp and an average standard deviation (SD) of 2.84 g kg $^{-1}$ cp. Figure 1 highlights distributions for all environments combined (1a), Blacksburg, VA 2018 and 2019 combined (1b), Warsaw, VA (1c), and Clayton, NC (1d). Blacksburg, Warsaw, and Clayton environments had means and SDs of 8.96, 12.32, and 5.88 g kg $^{-1}$ cp and 3.36, 1.73, and 2.61 g kg $^{-1}$ cp, respectively. Warsaw, VA exhibited significantly higher average Met than both other locations, while Blacksburg, VA also possessed significantly higher average Met than Clayton, NC. Samples grown in 2019 showed significantly higher Met content than 2018, but interaction variables could not be analyzed since locations and years were not orthogonal. Accessions were not shown to have a significant impact on Met content.

Population Structure

Through DAPC, 150 principal components that accounted for 78% of cumulative variance were retained, and with the smallest BIC, $k = 4$ was determined as the optimal number of clusters (Figure 2). Country of origin for accessions within each cluster were identified (Table 1). Cluster I (n = 76) contained 55 accessions (72.4%) that originated from China, 11 from Vietnam (14.5%), five from Japan (6.6%), three from Taiwan (3.9%), and one from Indonesia (1.3%). Cluster I also contained 52.6% of accessions from maturity group (MG) V. Cluster II (n = 62) contained 54 (87.1%), four (6.5%), two (3.2%), one (1.6%), and one (1.6%) accessions from China, Japan, the United States, Georgia, and South Korea, respectively, and 83.9% of those belonged to MG IV. Cluster

III (n = 47) contained 37 (78.7%) accessions from the United States, three (6.4%) from South Korea, two (4.3%) from Japan, and one (2.1% each) from Australia, Brazil, and Costa Rica. Cluster III also contained 78.7% of accessions from MG IV. Cluster IV (n = 126) contained 65 (51.6%), 15 (11.9%), 14 (11%), 11 (8.7%), and seven (5.6%) accessions from China, Japan, South Korea, the United States, and North Korea, respectively, as well as two (1.6% each) accessions from Georgia, Uganda, and Vietnam and one accession (0.8% each) from Brazil, India, Morocco, Nepal, Russia, and Taiwan. Within cluster IV, 77% of accessions belonged to MG IV. Clusters were not shown to have a significant effect on Met content, however the clusters proved useful in identifying genetically similar accessions that were stratified predominantly by geographic origin.

Genome-wide Associations

A total of 23 SNPs were identified as being associated with proteinogenic Met concentration (g kg $^{-1}$ cp) in soybean seed (Table 2). MLM and GLM models from 2018 environments displayed three SNPs (one SNP from each model) above the suggestive threshold (Figure 3), whereas MLM and GLM models from 2019 environments displayed 20 SNPs above the suggestive threshold (six from Blacksburg, VA, nine from Warsaw, VA, and five from a combined locations) (Figure 4). QQ plots for each model exhibited that Type I and Type II errors were accounted for sufficiently (Figures 3 and 4). Eight SNPs displayed significant associations [$-\log_{10}(P) > 4.91$]: ss715586112, ss715586120, ss715586126, ss715586203, ss715586204, ss715599541, ss715599547, and ss715625009. The remaining 15 SNPs displayed $-\log_{10}(P) > 4.16$ which was above the suggestive threshold: ss715585365, ss715586063, ss715586201, ss715589347, ss715589348, ss715589349, ss715590327, ss715593682, ss715593752, ss715625002,

ss715625007, ss715625012, ss715625013, and ss715625017. Chromosome (Chr) 3 contained the most associations (five significant, three suggestive), followed by Chr 16 (one significant, five suggestive), Chr 4 (three suggestive), Chr 6 (two suggestive), Chr 8 (two significant), Chr 5 (one suggestive), and Chr 12 (one suggestive). When including all environments, an MLM did not identify any SNPs above the significance or suggestive threshold.

Candidate Genes

A total of 22 candidate gene models from Wm82 were identified within 10 kb flanking regions of each significant SNP (Table 3). A number of gene models were found on three chromosomes: 13 on Chr 3 (Glyma.03g188100, Glyma.03g188200, Glyma.03g188300, Glyma.03g188400, Glyma.03g188900, Glyma.03g189000, Glyma.03g189100, Glyma.03g189700, Glyma.03g189800, Glyma.03g203900, Glyma.03g204000, Glyma.03g204100, and Glyma.03g204200), seven on Chr 8 (Glyma.08g177000, Glyma.08g177100, Glyma.08g177200, Glyma.08g177300, Glyma.08g177400, Glyma.08g177500, and Glyma.08g177600), and two on Chr 16 (Glyma.16g219800 and Glyma.16g219900). Candidate gene models belong to several protein families with numerous metabolic and biosynthesis implications. Of the 13 genes present on Chr 3, nine displayed moderate to high expression in reproductive tissues. Specifically, Glyma.03g188900, a ubiquitin-protein ligase, and Glyma.03g189800, a leucine-rich repeat (LRR) protein kinase, displayed high expression in all reproductive tissue and pods, respectively. On Chr 8, four out of seven genes had moderate to high expression in reproductive tissue, including Glyma.08g177000 a RING/U-box

superfamily protein. On Chr 16, Glyma.16g219800 displayed little to no expression in reproductive tissue, and Glyma.16g219900 did not had available expression data.

Genomic Selection

gBLUP through TASSEL estimated GEBVs using two different sets of SNPs: a complete set with 35,570 SNPs and a subset of 248 SNPs with some association with Met content ($-\log_{10}(P) > 3$). The correlation coefficient (r^2) between GEBVs and observed values varied throughout environments, but the subset of 248 SNPs consistently outperformed the larger SNP set (Figure 5). Using the larger set, the average r^2 for 2018 Blacksburg, VA, 2018 Clayton, NC, 2019 Blacksburg, VA, and 2019 Warsaw, VA datasets was 0.27, 0.03, 0.08, and 0.14, respectively. Using the 248 SNP subset, the average r^2 for 2018 Blacksburg, VA, 2018 Clayton, NC, 2019 Blacksburg, VA, and 2019 Warsaw, VA datasets was 0.62, 0.45, 0.48, and 0.48, respectively. When averaging Met content across all environments, prediction accuracy remained consistent, 0.05 and 0.41 average r^2 for the complete set and subset, respectively. T-tests comparing r^2 between SNP sets within environments identified that accuracy when using the subset was significantly higher across all environments ($P < 0.01$).

Discussion

Soybean protein content and amino acid profiles are critical objectives for plant breeders. For this reason, many resources have been allocated to unlock genomic controls for these traits. As suggested by Lee et al. (2019) and Jarquin et al. (2016), utilizing the high-density marker set from the SoySNP50K repository with environmentally suitable accessions in replicated, multi-location trials is a powerful method for unlocking genetic potential. In this study, we identified novel associations for proteinogenic Met content (g kg $^{-}$1 cp) in soybean seeds using accessions from MG IV and V that complements current genomic knowledge. Furthermore, we discovered that GS with a subset of significantly associated SNPs improved the genomic prediction accuracy for Met.

Previous studies have identified genomic associations with Met content on chromosomes 1, 2, 6, 7, 9, 10, 11, 12, 13, 14, 15, 17, 18, and 20 (Panthee et al., 2006a; Fallen et al., 2013; Kastoori Ramamurthy et al., 2014; Warrington et al., 2015; Zhang et al., 2018; Lee et al., 2019; Malle et al., 2020). Although our study did not identify these same genetic regions, ss715593752 on Chr 6 was within 220 kb of a QTL from Warrington et al. (2015) and a suggested SNP from Lee et al. (2019). Additionally, ss715593682 is within 6,000 kb of a SNP identified by Zhang et al. (2018). Through GWAS, we identified 23 novel SNP associations for proteinogenic Met content that were not recurrent across environment, which is consistent with previous research (McClure et al., 2017; Lee et al., 2019). This suggests further research is needed to understand GxE interactions for amino acid profile improvements in soybean due to their complexity.

Our analyzes identified associations greater in number and significance from the 2019 dataset when compared to 2018 measurements. This is likely caused by substantial

differences between Met concentrations between years. In Figure 1, the histogram for Warsaw, VA displays an expected frequency distribution for Met content, whereas other distributions exhibit numerous measurements below expected levels as a result of included 2018 data. Soybeans harvested from both locations in 2018 exhibited poorer seed quality likely as a function of higher than normal precipitation rates late in the growing season and delayed harvest. Rainfall has been shown to have a negative correlation with protein content (Kumar et al., 2006) and delayed harvest dates decrease concentrations of seed components (Jaureguy et al., 2013). These factors combined with higher disease rates, due to increased moisture, likely impacted the proteinogenic Met content. The three SNP associations from 2018 data exhibited a $-\log_{10}(P)$ greater than the suggestive threshold, but not the significance threshold. Although, ss715590327 (suggested from combined 2018 environments) was within 10 kb of Glyma.05g104400, a gene model involved in peptidyl-amino acid modification.

The 20 SNPs identified from our 2019 datasets provide superior evidence for associations to Met concentrations. The strongest associations occurred on Chr 3 with a set of four SNPs (ss715586063, ss715586112, ss715586120, and ss715586126) within a distance of 710 kb and another set of three SNPs (ss715586201, ss715586203, and ss715586204) within a distance of 20 kb. Within immediate proximity to the former set, nine gene models of relevant protein functions are present with ss715586126 being inside the coding region of Glyma.03g18980, a leucine-rich repeat protein kinase family protein that is highly expressed in pod walls. The latter set is close to four gene models including Glyma.03g204000, a Mal d 1-associated protein expressed highly in the root system and

moderately in pods and developing seeds, where ss715586203 is within its coding sequence.

While only suggestive associations, two SNPs on Chr 6 are within a 300 kb distance, and ss715593682 is part of the coding region for a S-adenosyl-L-methionine-dependent methyltransferase, Glyma.06g193300. The two significant SNPs found on Chr 8 (ss715599541 and ss715599547) are within 31 kb of each other and are proximal to seven various genes. Interestingly, ss715599541 is a part of the 3' untranslated region of Glyma.08g177100, a gene model with unknown function. Chr 16 contains one significant SNP association (ss715625009) that is flanked by five other suggestive associations, all within a 124 kb region. Within this region, ss715625012 can be found in the coding sequence of Glyma.16g220200, a serine/threonine protein kinase.

When our results are combined with previously identified QTLs, genomic regions impacting Met concentration in soybean seeds can be found on all chromosomes except Chr 19. This creates a complicated framework for increasing Met content through marker-assisted selection (MAS), transgenic, or genome editing approaches. Amir et al. (2019) summarized current efforts at biofortification of Met in plant seeds through gene regulation and found that most attempts failed to increase Met in a synergistic manner. More specifically, some researchers have incorporated cystathionine γ-synthase genes from *Arabidopsis thaliana* into soybean; Song et al. (2013) found an increase in general Met content, whereas Hanafy et al. (2013) saw increased soluble Met but not total Met in seeds. In *Arabidopsis thaliana,* Girija et al. (2020) discovered that Met protein residues, unsoluble Met production was the limiting factor for final Met content in seeds.

In breeding applications, our study suggests that GS may be a useful tool for selecting varieties with increased Met content. GS success is mainly determined by prediction accuracy (Duhnen et al., 2017) and impacted by many variables, including marker density. While high-density marker sets are typically ideal for utilizing genome-wide data, subsets of significant SNPs have been found to perform equal to or better than large SNP collections (Zhang et al., 2016; Qin et al., 2019). Qin et al. (2019) specifically identified improved genomic prediction for soybean amino acid content using a subset of 231 SNPs. Our results showed similar improvement in prediction accuracies with a subset of 248 SNPs. In 2018 Clayton, NC, both 2019 environments, and using average Met content, GS had average accuracy values between 0.41-0.48. This could prove useful to breeders and may complement the use significant SNPs from the 2019 dataset with MAS. However, when using the 2018 Blacksburg, VA dataset, predictive accuracy reached and average of 0.62. Considering the single suggestive SNP identified through GWAS for this location, GS appears to provide greater utility.

In summary, this project included a GWAS that not only identified many SNPs associated with Met content but also characterized several genomic regions that appear relevant. Within these regions, numerous gene models are present and their expression may correlate to the desired trait. GS was also evaluated as a potential method for selecting soybean lines with higher Met content. GS appears to be useful in certain environments with a subset of SNPs and could complement or outperform MAS. However, GxE limitations are still present and may impact which genes are influencing the final Met concentrations. This will require further research to elucidate genomic control of Met concentrations in soybean seed.

References

AOAC International. Official Methods of Analysis. http://www.eoma.aoac.org/methods/info.asp?ID=32703 (accessed 29 September 2021).

Amir, R., H. Cohen, and Y. Hacham. 2019. Revisiting the attempts to fortify methionine content in plant seeds. Journal of Experimental Botany 70(16): 4105–4114. doi: 10.1093/jxb/erz134.

Ashburner, M., C.A. Ball, J.A. Blake, D. Botstein, H. Butler, et al. 2000. Gene Ontology: tool for the unification of biology. Nat Genet 25(1): 25–29. doi: 10.1038/75556.

Berardini, T.Z., L. Reiser, D. Li, Y. Mezheritsky, R. Muller, et al. 2015. The Arabidopsis information resource: Making and mining the "gold standard" annotated reference plant genome. Genesis 53(8): 474–485. doi: 10.1002/dvg.22877.

Berry, T.H., D.E. Becker, O.G. Rasmussen, A.H. Jensen, and H.W. Norton. 1962. The Limiting Amino Acids in Soybean Protein. Journal of Animal Science 21(3): 558–561. doi: 10.2527/jas1962.213558x.

Boisen, S., T. Hvelplund, and M.R. Weisbjerg. 2000. Ideal amino acid profiles as a basis for feed protein evaluation. Livestock Production Science 64(2): 239–251. doi: 10.1016/S0301-6226(99)00146-3.

Bonato, M.A., N.K. Sakomura, J.C. Siqueira, J.B.K. Fernandes, and R.M. Gous. 2011. Maintenance requirements for methionine and cysteine, and threonine for poultry. South African Journal of Animal Science 41(3): 209-222–222. doi: 10.4314/sajas.v41i3.3.

Bradbury, P.J., Z. Zhang, D.E. Kroon, T.M. Casstevens, Y. Ramdoss, et al. 2007. TASSEL: software for association mapping of complex traits in diverse samples. Bioinformatics 23(19): 2633–2635. doi: 10.1093/bioinformatics/btm308.

Brosnan, J.T., M.E. Brosnan, R.F.P. Bertolo, and J.A. Brunton. 2007. Methionine: A metabolically unique amino acid. Livestock Science 112(1): 2–7. doi: 10.1016/j.livsci.2007.07.005.

Bunchasak, C. 2009. Role of Dietary Methionine in Poultry Production. The Journal of Poultry Science 46(3): 169–179. doi: 10.2141/jpsa.46.169.

Buttery, P.J., and J.P.F. D'Mello. 1994. Amino Acid Metabolism in Farm Animals: An Overview. Amino Acids in Farm Animal Nutrition. CAB International, Wallingford, United Kingdom. p. 1–10

Conde-Aguilera, J.A., C. Cobo-Ortega, S. Tesseraud, M. Lessire, Y. Mercier, et al. 2013. Changes in body composition in broilers by a sulfur amino acid deficiency during growth. Poult. Sci. 92(5): 1266–1275. doi: 10.3382/ps.2012-02796.

Duhnen, A., A. Gras, S. Teyssèdre, M. Romestant, B. Claustres, et al. 2017. Genomic Selection for Yield and Seed Protein Content in Soybean: A Study of Breeding Program Data and Assessment of Prediction Accuracy. Crop Science 57(3): 1325–1337. doi: 10.2135/cropsci2016.06.0496.

Endelman, J.B., and J.-L. Jannink. 2012. Shrinkage Estimation of the Realized Relationship Matrix. G3 (Bethesda) 2(11): 1405–1413. doi: 10.1534/g3.112.004259.

Fallen, B., C. Hatcher, F. Allen, D. Kopsell, A. Saxton, et al. 2013. Soybean Seed Amino Acid Content QTL Detected Using the Universal Soy Linkage Panel 1.0 with 1,536 SNPs. Agronomy & Horticulture -- Faculty Publications. https://digitalcommons.unl.edu/agronomyfacpub/807.

Fernandez, S.R., S. Aoyagi, Y. Han, C.M. Parsons, and D.H. Baker. 1994. Limiting order of amino acids in corn and soybean meal for growth of the chick. Poult. Sci. 73(12): 1887–1896. doi: 10.3382/ps.0731887.

Gene Ontology Consortium. 2021. The Gene Ontology resource: enriching a GOld mine. Nucleic Acids Res 49(D1): D325–D334. doi: 10.1093/nar/gkaa1113.

Girija, A., D. Shotan, Y. Hacham, and R. Amir. 2020. The Level of Methionine Residues in Storage Proteins Is the Main Limiting Factor of Protein-Bound-Methionine Accumulation in Arabidopsis Seeds. Front Plant Sci 11: 1136. doi: 10.3389/fpls.2020.01136.

Hanafy, M.S., S.M. Rahman, Y. Nakamoto, T. Fujiwara, S. Naito, et al. 2013. Differential response of methionine metabolism in two grain legumes, soybean and azuki bean, expressing a mutated form of Arabidopsis cystathionine γ-synthase. Journal of Plant Physiology 170(3): 338–345. doi: 10.1016/j.jplph.2012.10.018.

Hwang, E.-Y., Q. Song, G. Jia, J.E. Specht, D.L. Hyten, et al. 2014. A genome-wide association study of seed protein and oil content in soybean. BMC Genomics 15(1): 1. doi: 10.1186/1471-2164-15-1.

Hymowitz, T., and J.R. Harlan. 1983. Introduction of soybean to North America by Samuel Bowen in 1765. Econ Bot 37(4): 371–379. doi: 10.1007/BF02904196.

Jarquin, D., J. Specht, and A. Lorenz. 2016. Prospects of Genomic Prediction in the USDA Soybean Germplasm Collection: Historical Data Creates Robust Models for Enhancing Selection of Accessions. G3: Genes, Genomes, Genetics 6(8): 2329–2341. doi: 10.1534/g3.116.031443.

Jaureguy, L.M., F.L. Rodriguez, L. Zhang, P. Chen, K. Brye, et al. 2013. Planting Date and Delayed Harvest Effects on Soybean Seed Composition. Crop Science 53(5): 2162–2175. doi: 10.2135/cropsci2012.12.0683.

Jombart, T. 2008. adegenet: a R package for the multivariate analysis of genetic markers. Bioinformatics 24(11): 1403–1405. doi: 10.1093/bioinformatics/btn129.

Jombart, T., S. Devillard, and F. Balloux. 2010. Discriminant analysis of principal components: a new method for the analysis of genetically structured populations. BMC Genetics 11(1): 94. doi: 10.1186/1471-2156-11-94.

Kastoori Ramamurthy, R., J. Jedlicka, G.L. Graef, and B.M. Waters. 2014. Identification of new QTLs for seed mineral, cysteine, and methionine concentrations in soybean [Glycine max (L.) Merr.]. Mol Breeding 34(2): 431–445. doi: 10.1007/s11032-014-0045-z.

Kuiken, K.A., and M. Lyman. 1949. Essential amino acid composition of soybean meals prepared from twenty strains of soybeans. J Biol Chem (177): 29–36.

Kumar, V., A. Rani, S. Solanki, and S.M. Hussain. 2006. Influence of growing environment on the biochemical composition and physical characteristics of soybean seed. Journal of food composition and analysis : an official publication of the United Nations University, International Network of Food Data Systems. http://dx.doi.org/10.1016/j.jfca.2005.06.005 (accessed 19 September 2021).

Lee, S., K. Van, M. Sung, R. Nelson, J. LaMantia, et al. 2019. Genome-wide association study of seed protein, oil and amino acid contents in soybean from maturity groups I to IV. Theor Appl Genet 132(6): 1639–1659. doi: 10.1007/s00122-019-03304-5.

Li, J., and L. Ji. 2005. Adjusting multiple testing in multilocus analyses using the eigenvalues of a correlation matrix. Heredity 95(3): 221–227. doi: 10.1038/sj.hdy.6800717.

Li, X., R. Tian, S. Kamala, H. Du, W. Li, et al. 2018. Identification and verification of pleiotropic QTL controlling multiple amino acid contents in soybean seed. Euphytica 214(6): 93. doi: 10.1007/s10681-018-2170-y.

Li, D., X. Zhao, Y. Han, W. Li, and F. Xie. 2019. Genome-wide association mapping for seed protein and oil contents using a large panel of soybean accessions. Genomics 111(1): 90–95. doi: 10.1016/j.ygeno.2018.01.004.

Ma, Y., J.C. Reif, Y. Jiang, Z. Wen, D. Wang, et al. 2016. Potential of marker selection to increase prediction accuracy of genomic selection in soybean (Glycine max L.). Mol Breeding 36(8): 113. doi: 10.1007/s11032-016-0504-9.

Malle, S., M. Eskandari, M. Morrison, and F. Belzile. 2020. Genome-wide association identifies several QTLs controlling cysteine and methionine content in soybean seed including some promising candidate genes. Sci Rep 10(1): 21812. doi: 10.1038/s41598-020-78907-w.

Matei, G., L.G. Woyann, A.S. Milioli, I. de Bem Oliveira, A.D. Zdziarski, et al. 2018. Genomic selection in soybean: accuracy and time gain in relation to phenotypic selection. Mol Breeding 38(9): 117. doi: 10.1007/s11032-018-0872-4.

McClure, T., J.-C. Cocuron, V. Osmark, L.K. McHale, and A.P. Alonso. 2017. Impact of Environment on the Biomass Composition of Soybean (Glycine max) seeds. J. Agric. Food Chem. 65(32): 6753–6761. doi: 10.1021/acs.jafc.7b01457.

Meuwissen, T.H., B.J. Hayes, and M.E. Goddard. 2001. Prediction of total genetic value using genome-wide dense marker maps. Genetics 157(4): 1819–1829.

Mi, H., A. Muruganujan, and P.D. Thomas. 2013. PANTHER in 2013: modeling the evolution of gene function, and other gene attributes, in the context of phylogenetic trees. Nucleic Acids Res 41(Database issue): D377-386. doi: 10.1093/nar/gks1118.

Neubauer, C., and H. Landecker. 2021. A planetary health perspective on synthetic methionine. The Lancet Planetary Health 5(8): e560–e569. doi: 10.1016/S2542-5196(21)00138-8.

Panthee, D., and V.R. Pantalone. 2006. Registration of Soybean Germplasm Lines TN03–350 and TN04–5321 with Improved Protein Concentration and Quality. Crop Science - CROP SCI 46. doi: 10.2135/cropsci2005.11.0437.

Panthee, D.R., V.R. Pantalone, C.E. Sams, A.M. Saxton, D.R. West, et al. 2006a. Quantitative trait loci controlling sulfur containing amino acids, methionine and cysteine, in soybean seeds. Theor Appl Genet 112(3): 546–553. doi: 10.1007/s00122-005-0161-6.

Panthee, D.R., V.R. Pantalone, A.M. Saxton, D.R. West, and C.E. Sams. 2006b. Genomic regions associated with amino acid composition in soybean. Molecular Breeding: 79–89. doi: 10.1007/s11032-005-2519-5.

Panthee, D.R., V.R. Pantalone, D.R. West, A.M. Saxton, and C.E. Sams. 2005. Quantitative Trait Loci for Seed Protein and Oil Concentration, and Seed Size in Soybean. Crop Science 45(5): 2015–2022. doi: 10.2135/cropsci2004.0720.

Patil, G., R. Mian, T. Vuong, V. Pantalone, Q. Song, et al. 2017. Molecular mapping and genomics of soybean seed protein: a review and perspective for the future. Theor Appl Genet 130(10): 1975–1991. doi: 10.1007/s00122-017-2955-8.

Qin, J., A. Shi, Q. Song, S. Li, F. Wang, et al. 2019. Genome Wide Association Study and Genomic Selection of Amino Acid Concentrations in Soybean Seeds. Front. Plant Sci. 10. doi: 10.3389/fpls.2019.01445.

Qin, J., Q. Song, A. Shi, S. Li, M. Zhang, et al. 2017. Genome-wide association mapping of resistance to Phytophthora sojae in a soybean [Glycine max (L.) Merr.] germplasm panel from maturity groups IV and V. PLOS ONE 12(9): e0184613. doi: 10.1371/journal.pone.0184613.

Ravelombola, W.S., J. Qin, A. Shi, L. Nice, Y. Bao, et al. 2019. Genome-wide association study and genomic selection for soybean chlorophyll content associated with soybean cyst nematode tolerance. BMC Genomics 20(1): 904. doi: 10.1186/s12864-019-6275-z.

Ravelombola, W.S., J. Qin, A. Shi, L. Nice, Y. Bao, et al. 2020. Genome-wide association study and genomic selection for tolerance of soybean biomass to soybean cyst nematode infestation. PLOS ONE 15(7): e0235089. doi: 10.1371/journal.pone.0235089.

Ravelombola, W., J. Qin, A. Shi, Q. Song, J. Yuan, et al. 2021. Genome-wide association study and genomic selection for yield and related traits in soybean. PLOS ONE 16(8): e0255761. doi: 10.1371/journal.pone.0255761.

Severin, A.J., J.L. Woody, Y.-T. Bolon, B. Joseph, B.W. Diers, et al. 2010. RNA-Seq Atlas of Glycine max: A guide to the soybean transcriptome. BMC Plant Biol 10(1): 160. doi: 10.1186/1471-2229-10-160.

Song, S., W. Hou, I. Godo, C. Wu, Y. Yu, et al. 2013. Soybean seeds expressing feedback-insensitive cystathionine γ-synthase exhibit a higher content of methionine. Journal of Experimental Botany 64(7): 1917–1926. doi: 10.1093/jxb/ert053.

Song, Q., D.L. Hyten, G. Jia, C.V. Quigley, E.W. Fickus, et al. 2015. Fingerprinting Soybean Germplasm and Its Utility in Genomic Research. G3: Genes, Genomes, Genetics 5(10): 1999–2006. doi: 10.1534/g3.115.019000.

Stewart-Brown, B.B., Q. Song, J.N. Vaughn, and Z. Li. 2019. Genomic Selection for Yield and Seed Composition Traits Within an Applied Soybean Breeding Program. G3 (Bethesda) 9(7): 2253–2265. doi: 10.1534/g3.118.200917.

The American Soybean Association. 2020. U.S. Soybean Meal: Use by Livestock. http://soystats.com/soybean-meal-u-s-use-by-livestock/ (accessed 14 August 2021).

Turner, S.D. 2014. qqman: an R package for visualizing GWAS results using Q-Q and manhattan plots.

Warrington, C.V., H. Abdel-Haleem, D.L. Hyten, P.B. Cregan, J.H. Orf, et al. 2015. QTL for seed protein and amino acids in the Benning × Danbaekkong soybean population. Theor Appl Genet 128(5): 839–850. doi: 10.1007/s00122-015-2474-4.

Willke, T. 2014. Methionine production—a critical review. Appl Microbiol Biotechnol 98(24): 9893–9914. doi: 10.1007/s00253-014-6156-y.

Wilson, R.F. 2004. Seed Composition. Soybeans: Improvement, Production, and Uses. 3rd ed. ASA, CSSA, SSSA, Madison, WI. p. 621–669

Wu, G. 2014. Dietary requirements of synthesizable amino acids by animals: a paradigm shift in protein nutrition. J Anim Sci Biotechnol 5(1): 34. doi: 10.1186/2049-1891-5-34.

Xie, D., Z. Dai, Z. Yang, J. Sun, D. Zhao, et al. 2018. Genome-Wide Association Study Identifying Candidate Genes Influencing Important Agronomic Traits of Flax (Linum usitatissimum L.) Using SLAF-seq. Frontiers in Plant Science 8: 2232. doi: 10.3389/fpls.2017.02232.

Yu, J., G. Pressoir, W.H. Briggs, I. Vroh Bi, M. Yamasaki, et al. 2006. A unified mixed-model method for association mapping that accounts for multiple levels of relatedness. Nat Genet 38(2): 203–208. doi: 10.1038/ng1702.

Zargar, S., B. Raatz, H. Sonah, MuslimaNazir, J. Bhat, et al. 2015. Recent advances in molecular marker techniques: Insight into QTL mapping, GWAS and genomic selection in plants. Journal of Crop Science and Biotechnology 18: 293–308. doi: 10.1007/s12892-015-0037-5.

Zhang, J., Q. Song, P.B. Cregan, and G.-L. Jiang. 2016. Genome-wide association study, genomic prediction and marker-assisted selection for seed weight in soybean (Glycine max). Theoretical and Applied Genetics 1(129): 117–130. doi: 10.1007/s00122-015-2614-x.

Zhang, J., X. Wang, Y. Lu, S.J. Bhusal, Q. Song, et al. 2018. Genome-wide Scan for Seed Composition Provides Insights into Soybean Quality Improvement and the Impacts of Domestication and Breeding. Molecular Plant 11(3): 460–472. doi: 10.1016/j.molp.2017.12.016.

Tables and Figures

Table 1. Countries of origin and maturity groups (MG) for clustered accessions as determined by discriminant analysis of principal components (DAPC).

	Cluster I (n = 76)		Cluster II (n = 62)		Cluster III (n = 47)		Cluster IV (n = 126)		
	Count	%	Count	%	Count	%	Count	%	Total
Australia	-	-	-	-	1	2.1	-		1
Brazil	-	-	-	-	1	2.1	1	0.8	2
China	55	72.4	54	87.1	-	-	65	51.6	174
Costa Rica	-	-	-	-	1	2.1	-		1
Georgia	-	-	1	1.6	-	-	2	1.6	3
India	-	-	-	-	-	-	1	0.8	1
Indonesia	1	1.3	-	-	-	-	-		1
Japan	5	6.6	4	6.5	2	4.3	15	11.9	26
Morocco	-	-	-	-	-	-	1	0.8	1
Nepal	-	-	-	-	-	-	1	0.8	1
North Korea	-	-	-	-	-	-	7	5.6	7
Russia	-	-	-	-	-	-	1	0.8	1
South Korea	-	-	1	1.6	3	6.4	14	11	18
Taiwan	3	3.9	-	-	-	-	1	0.8	4
Uganda	-	-	-	-	-	-	2	1.6	2
United States	-	-	2	3.2	37	78.7	11	8.7	50
Vietnam	11	14.5	-	-	-	-	2	1.6	13
Unknown	1	1.3	-	-	2	4.3	2	1.6	5
MG IV	36	47.4	52	83.9	37	78.7	97	77	222
MG V	40	52.6	10	16.1	10	21.3	29	23	89

Table 2. Significant SNPs on chromosomes 3, 4, 5, 6, 8, 12, and 16 associated with Met content (g kg $^{-1}$ cp) in soybean seeds.

Chr	Genomic Location	SNP (position)	Wm82 Allele[a]	Alter-native Allele	Environments[d]					
					2018 BB	2018 CL	2018 Combined	2019 BB	2019 W	2019 Combined
					---------- -log$_{10}$(P) ----------					
3	Intergenic	ss715585365 (33765404)	T	G	NS[b]	4.29*	NS	NS	NS	NS
	Intergenic	ss715586063 (39357229)	C	T	NS	NS	NS	4.60*	NS	NS
	Intergenic	ss715586112 (39946374)	A	G	NS	NS	NS	5.82**	NS	NS
	Intergenic	ss715586120 (40006278)	A	G	NS	NS	NS	5.16**	NS	NS
	Coding sequence	ss715586126 (40062294)	T	G	NS	NS	NS	5.57**	NS	NS
	Intergenic	ss715586201 (41217558)	A	G	NS	NS	NS	NS	NS	4.37*
	Coding sequence	ss715586203 (41228895)	G	T	NS	NS	NS	NS	NS	5.33**
	Intergenic	ss715586204 (41236923)	G	A	NS	NS	NS	NS	NS	5.11**
4	Coding sequence	ss715589347 (8089953)	T	C	NS	NS	NS	NS	4.27*	NS
	Intron	ss715589348 (8091107)	G	A	NS	NS	NS	NS	4.33*	NS
	Coding sequence	ss715589349 (8095691)	C	T	NS	NS	NS	NS	4.33*	NS
5	Intergenic	ss715590327 (27762168)	A	G	NS	NS	4.17*	NS	NS	NS
6	Coding sequence	ss715593682 (17154269)	G	A	NS	NS	NS	NS	NS	4.39*
	Intergenic	ss715593752 (17453327)	C	T	NS	NS	NS	NS	NS	4.20*
8	3' UTR[c]	ss715599541 (14196322)	T	C	NS	NS	NS	4.92**	NS	NS
	Intergenic	ss715599547 (14226774)	G	A	NS	NS	NS	5.81**	NS	NS
12	Intergenic	ss715613175 (5433032)	T	G	4.22*	NS	NS	NS	NS	NS
16	Intron	ss715625002 (37660795)	A	C	NS	NS	NS	NS	4.78*	NS
	Intergenic	ss715625007 (37701598)	T	G	NS	NS	NS	NS	4.38*	NS
	Intergenic	ss715625009 (37712387)	T	C	NS	NS	NS	NS	5.05**	NS
	Coding sequence	ss715625012 (37737235)	C	T	NS	NS	NS	NS	4.71*	NS
	Intergenic	ss715625013 (37753573)	T	C	NS	NS	NS	NS	4.74*	NS
	Intergenic	ss715625017 (37784014)	T	C	NS	NS	NS	NS	4.78*	NS

** significance threshold (5%), * suggestive threshold (25%)
[a]Williams 82
[b]not significant
[c]3 prime untranslated region
[d]Blacksburg, VA (BB), Clayton, NC (CL), Warsaw, VA (W)

Table 3. Candidate gene models and descriptions within 10 kb flanking regions of significantly associated SNPs using Wm82.a2.v1.

Chr	SNP	Candidate Genes	Gene Function Description[a]	Expression in soybean reproductive tissue[b]
3	ss715586112	Glyma.03g188100	Modifier of rudimentary protein	High expression in flowers
		Glyma.03g188200	Nucleic acid binding	NA
		Glyma.03g188300	Pollen Ole e 1 allergen and extensin family protein	Little to no expression in reproductive tissue
		Glyma.03g188400	Eukaryotic aspartyl protease family protein	Moderate to high expression in seeds and pods
	ss715586120	Glyma.03g188900	Ubiquitin-protein ligase 7	High expression in flowers, pods, and seeds
		Glyma.03g189000	Pentatricopeptide repeat (PPR) superfamily protein	Moderate to high expression in flowers, pods, and seeds
		Glyma.03g189100	Exostosin family protein	Moderate to high expression in seeds
	ss715586126	Glyma.03g189700	Pyruvate kinase family protein	Moderate to high expression in seeds
		Glyma.03g189800	Leucine-rich repeat (LRR) protein kinase family protein	High expression in pods
	ss715586203	Glyma.03g203900	Polyketide cyclase/dehydrase/lipid transport superfamily protein	NA
		Glyma.03g204000	Mal d 1-associated protein	Moderate expression in flowers, pods, and seeds
		Glyma.03g204100	Calmodulin-domain protein kinase cdpk isoform 2	Moderate to high expression in pods
	ss715586204	Glyma.03g204200	TPX2 (targeting protein for Xklp2) protein family	Little to no expression in reproductive tissue
8	ss715599541	Glyma.08g177000	RING/U-box superfamily protein	High expression in flower and pods
		Glyma.08g177100	NA	Little to no expression in reproductive tissue
		Glyma.08g177200	Arabinogalactan protein 1	NA
		Glyma.08g177300	GTP cyclohydrolase II	Little to no expression in reproductive tissue
	ss715599547	Glyma.08g177400	Dicarboxylate transport 2.1	Moderate expression in pods and seeds
		Glyma.08g177500	Pyrimidine 2	Moderate expression in flowers
		Glyma.08g177600	Centrin2	High expression in flowers; moderate expression in pods
16	ss715625009	Glyma.16g219800	WRKY DNA-binding protein 70	Little to no expression in reproductive tissue
		Glyma.16g219900	B-block binding subunit of TFIIIC	NA

[a]as described in TAIR, PANTHER, or GO annotation

[b]Soybean flowers, seeds, and pods. Detailed expression profiles can be found in Severin et al. (2010)

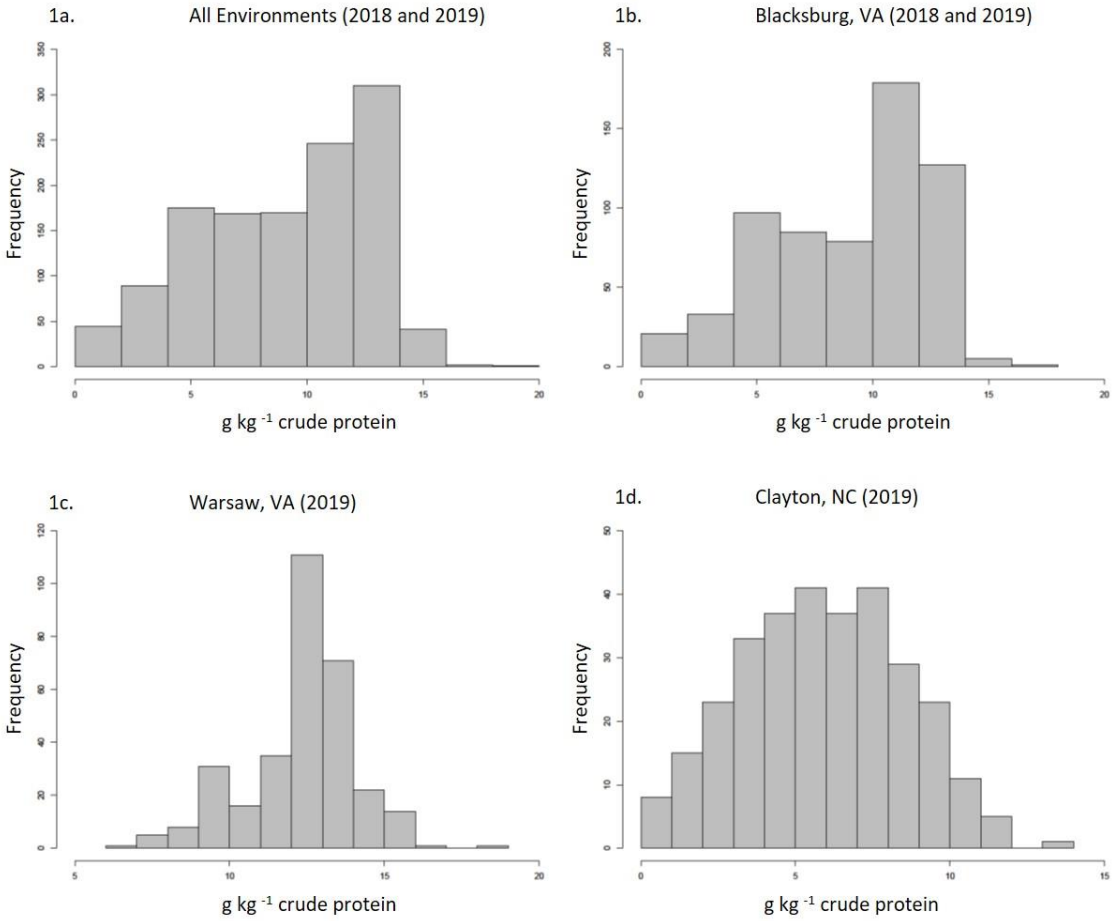

Figure 1. Frequency distributions displaying proteinogenic Met concentrations collected from all environments (1a), Blacksburg, VA (1b), Warsaw, VA (1c), and Clayton, NC (1d).

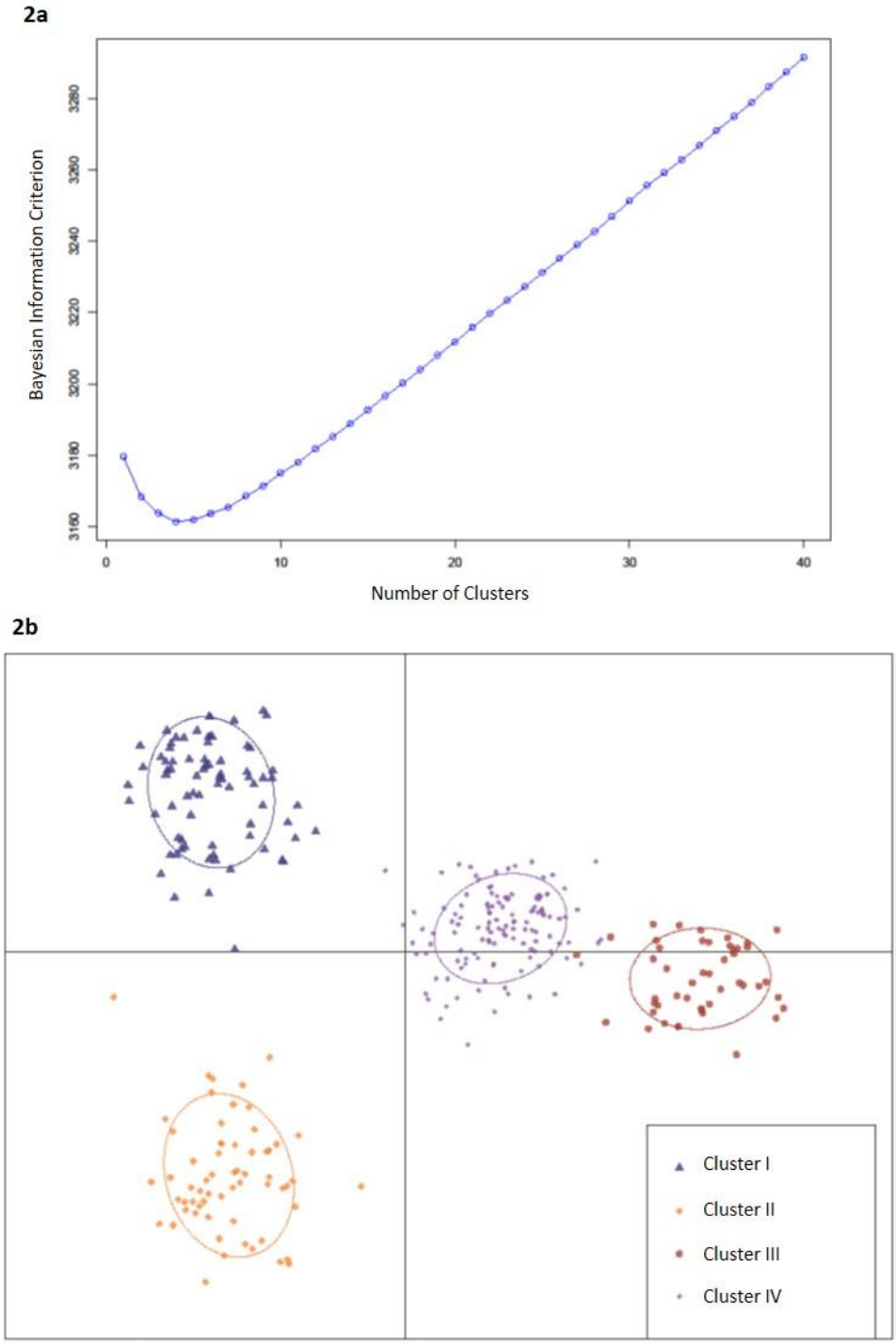

Figure 2. (2a) Bayesian information criterion for selecting the optimal number of clusters. (2b) A scatter plot depicting the 4 clusters (k = 4) identified as likely subpopulations within the 311 accessions: cluster I (blue triangle, n = 76), cluster II (gold diamonds, n = 62), cluster III (large red circles, n = 47), cluster IV (small purple circles, n = 126).

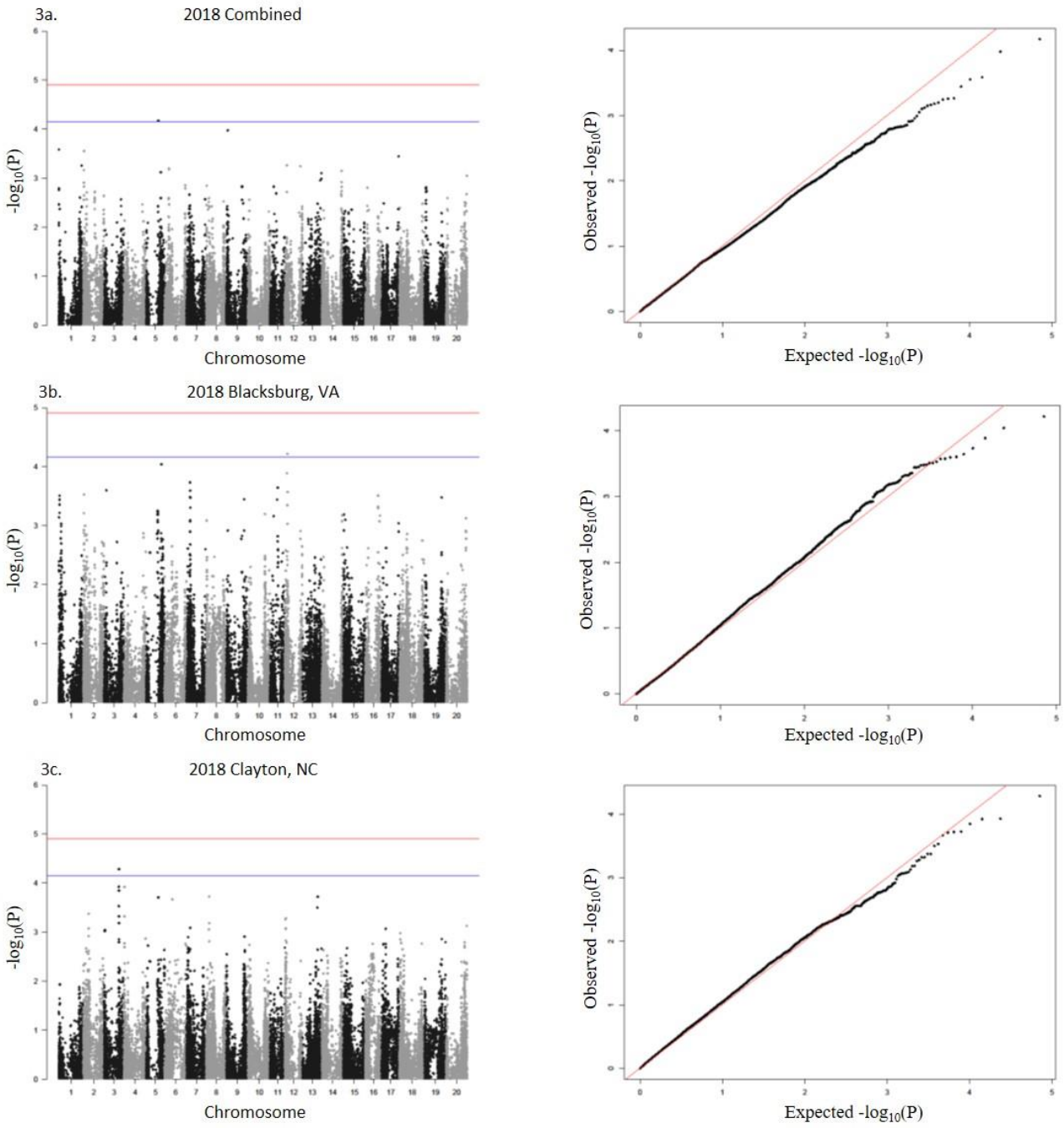

Figure 3. SNP associations for 2018 environments (3a. combined, 3b. Blacksburg, VA, 3c. Clayton, NC) are displayed in Manhattan plots with chromosomes in alternating colors, significance threshold -$\log_{10}$(P) > 4.91 and suggestive threshold -$\log_{10}$(P) > 4.16. Each respective QQ plot displays observed -$\log_{10}$(P) against expected -$\log_{10}$(P).

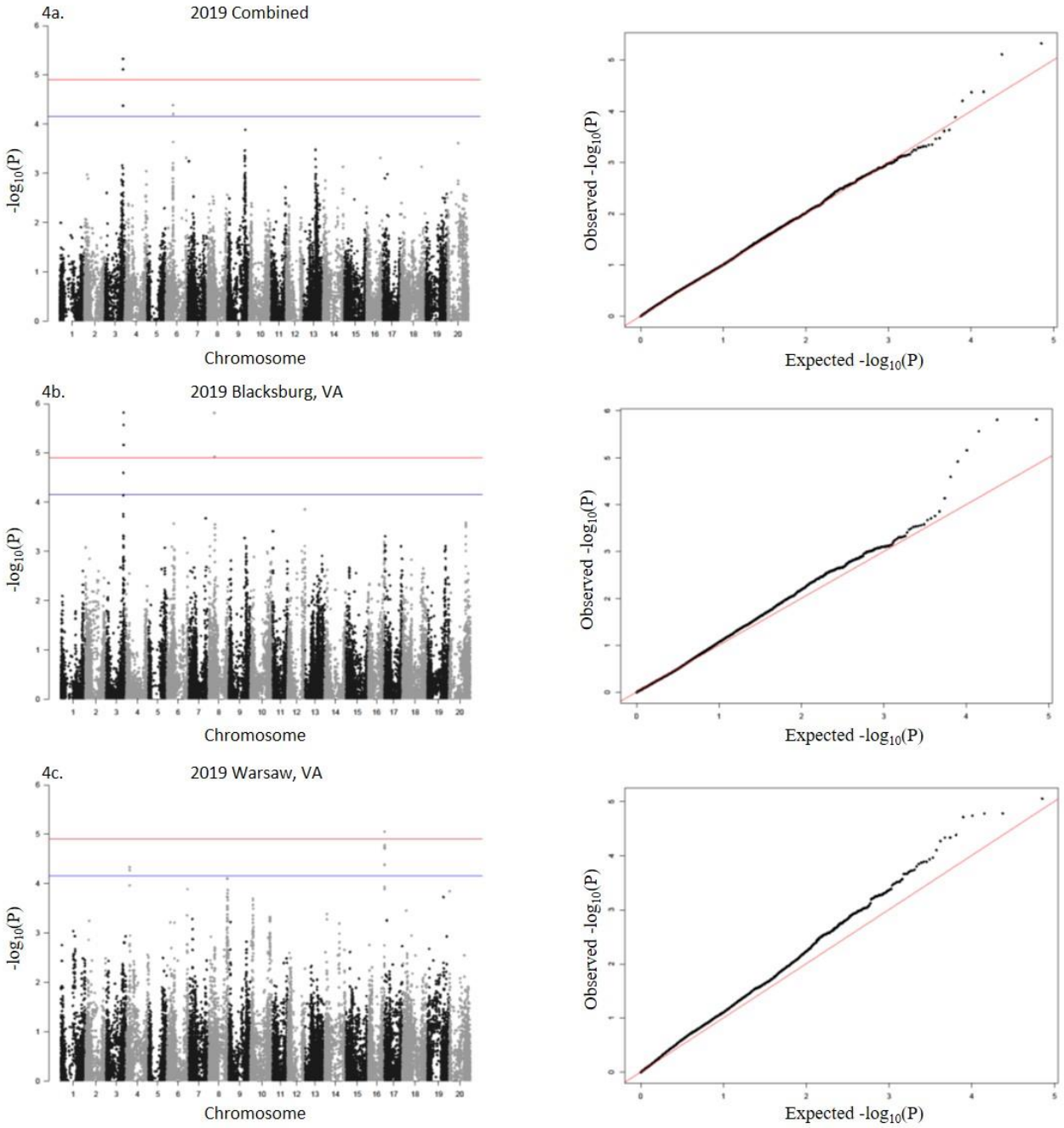

Figure 4. SNP associations for 2019 environments (4a. combined, 4b. Blacksburg, VA, 4c. Warsaw, VA) are displayed in Manhattan plots with chromosomes in alternating colors, significance threshold -$\log_{10}$(P) > 4.91 and suggestive threshold -$\log_{10}$(P) > 4.16. Each respective QQ plot displays observed -$\log_{10}$(P) against expected -$\log_{10}$(P).

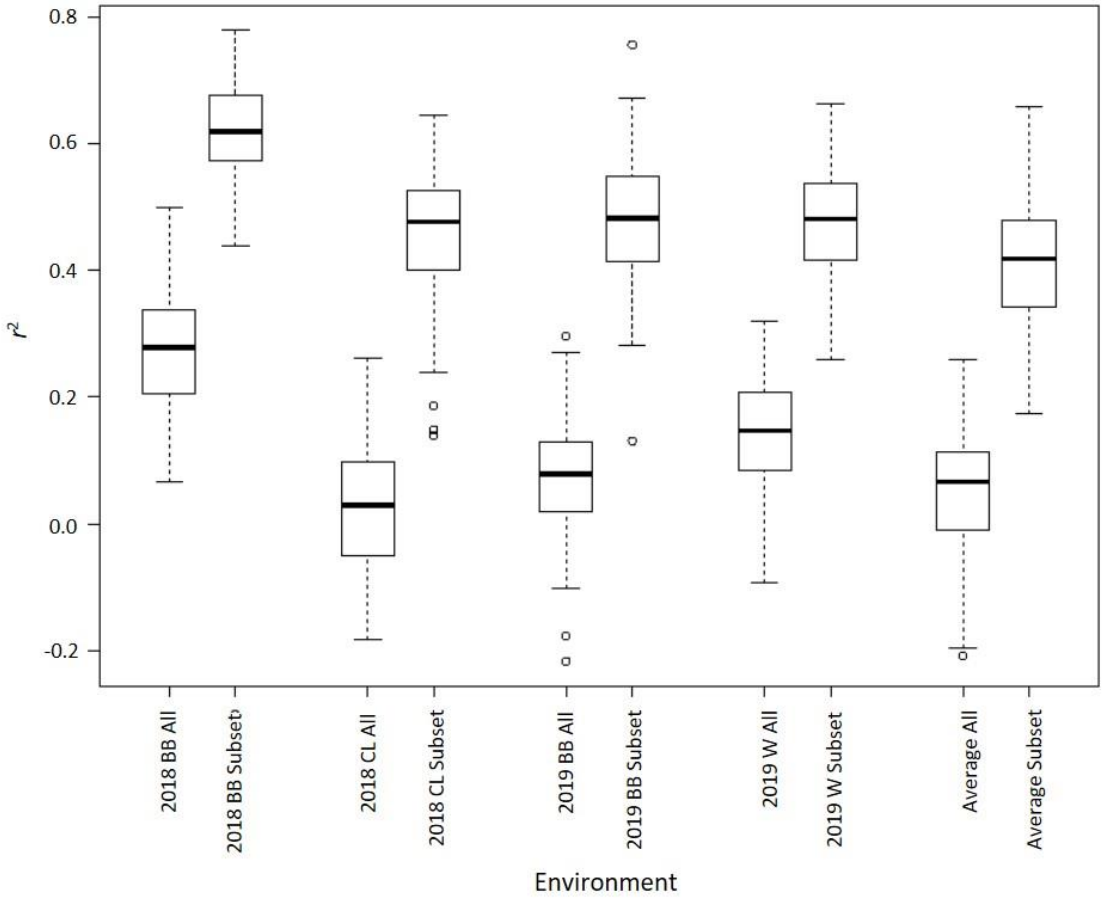

Figure 5. Boxplots displaying 100 r^2 values (k =5, 20 iterations) for GS models using 35,570 SNPs (All) and 248 SNPs (Subset) across environments (BB = Blacksburg, VA; CL = Clayton, NC; W = Warsaw, VA; Average = Mean Met across all environments).

Chapter 3: Comparison of Genomic and Phenomic Prediction Accuracy for Soybean Test Weight

William M. Singer[1], Zachary Shea[1], and Bo Zhang[1]

[1]School of Plant and Environmental Sciences, Virginia Tech

This chapter was submitted to Processes on September 29th, 2021.

Abstract

Soybean [*Glycine max (L.) Merr.*] is a valuable, global crop with many end-use purposes due to its seed composition. Test weight, a measure of bulk density, is directly linked to end-use efficiency as an important indicator of seed quality. Test weight in soybean has historically been overlooked as a trait of interest by plant breeders and is difficult to measure in early breeding generations due to the limited seed amount. The objective of this study was to compare genomic and phenomic prediction models for test weight of soybean seeds in order to accelerate its breeding selection. We planted 129 soybean accessions from maturity groups IV and V in two locations in 2019 and 2020. Genomic prediction was evaluated using 35,454 single nucleotide polymorphisms, k-fold cross validation, and with population structure determined through discriminant analysis of principal components. Phenomic prediction was performed using near-infrared reflectance spectral data as predictors with a standard least squares approach and with 80% training and 20% validation populations. Average phenomic prediction accuracy consistently outperformed genomic prediction with accuracies of 0.75 and 0.07-0.31, respectively. Phenomic prediction was also successful in cross validations by environment with an accuracy range of 0.16-0.59. These results suggest that phenomic prediction could be a useful selection tool for improving test weight in soybean.

Introduction

Soybean [*Glycine max (L.) Merr.*] is a valuable, global crop with a myriad of end-use purposes due to its seed composition. The protein and oil components within soybean seeds make up of 40% and 20%, respectively (Wilson, 2004). A variety of processing methods can unlock the value of soybean seeds by separating the protein and oil to create value-added products such as protein meal, cooking and industrial oils, and lecithin. Test weight, a measure of bulk density and an important indicator of seed quality, is intricately linked to this final utility. According to the U.S. Soybean Export Council, soybean test weight is graded No. 1, 2, 3, and 4 representing 72.1, 69.5, 66.9, and 63.1 kg hL^{-1} (U.S. Soybean Export Council, 2015). Aside from test weight's influence on seed quality, discounts may directly affect farmer profits. Price discounts from $0.005 to $0.02 per hectoliter may be applied for test weights beneath grade No. 2, and rejection can occur with test weights below No. 4 (Mississippi Soybean Promotion Board,). Additionally, poor test weight negatively impacts transportation and processing efficiency.

In 2019 and 2020, the U.S. exported $16.5 billion (USD) worth of whole soybeans with China accounting for $5.85 billion (The American Soybean Association, 2020). The storage, transportation, and export of soybean is significantly impacted by test weight. Soybean with poor test weight will occupy the same volume with less mass than equivalent soybeans with better test weight. For example, 1.4 to 1.55 million kg of soybean with a test weight of 77.2 kg hL^{-1} can be transported by one barge, 16 rail cars, or 62 semi-truck trailers (McNeece et al., 2021; Soy Transportation Coalition,). However, the average test weight of U.S. soybean in 2018 was 73.1 kg hL^{-1} (Naeve and Miller-Garvin, 2019). This would result in a potential reduction 77,500 kg of soybean per unit of

transport, and a potential loss over $36,000 (assuming a reasonable price of $12.74 per bushel (Markets Insider). In a competitive global market, maximizing transportation efficiency is paramount.

While test weight is relatively simple trait to measure in soybean with many options for mimicking grain elevator measurements in breeding programs, a consistent seed yield is required for accurate results. We have identified 13 oz or 370 g as the minimum seed amount needed for determining test weight with a widely used GAC 2500-AGRI Grain Analysis Computer (DICKEY-john, Auburn, IL). This seed quantity is unrealistic in pre-breeding materials and unattainable in early breeding generations, resulting in that test weight can't be selected until progeny rows are developed. Therefore, there is a need to easily and accurately evaluate test weight for early selections to accelerate its breeding process.

While test weight in relation to genotypic control, agronomic practices, and seed composition has been studied heavily in other grain crops such as small grains, soybean test weight has historically been overlooked. Recently, two articles by Liu et al. and McNeece et al. have identified genetic variation among genotypes in soybean and reported correlations between test weight and yield, maturity, protein content, oil content, sucrose content, seed size, and seed quality (Liu et al., 2019; McNeece et al., 2021). Research in other crops highlight greater potential for breeding improvements in test weight. In wheat (*Triticum*), researchers have developed a foundation for genotypic effects of test weight (Ghaderi et al., 1971; Teich, 1984; Rharrabti et al., 2001; Jing et al., 2003; Kaya and Akcura, 2014), and test weight inheritance and recurrent selection have been studied in oat (*Avena sativa*) (Pixley and Frey, 1991; Klein et al., 1993).

Additionally, quantitative trait loci (QTL) have been identified for test weight in maize (*Zea mays*) (Ding et al., 2011).

High-throughput genotyping technology, especially the use of single nucleotide polymorphisms (SNPs) as molecular markers, has allowed modern plant breeders to utilize high-density genetic markers for variety development. Specifically, genomic prediction (GP) uses genetic information to predict traits, while genomic selection (GS) refers active selection of breeding lines based upon those predictions. Crossa et al. aptly reviewed recent progress for genomic selection in plant breeding and discussed methods and statistical models used in selection (Crossa et al., 2017). GS accuracy is usually represented as the r^2 correlation between predicted and observed values after datasets have been divided into training and validation populations. Regarding test weight, GS has shown variable accuracy in rice (*Oryza sativa*) and small grains. Prediction accuracy was shown to increase as population size increased in rice but remained below 0.5 (Isidro et al., 2015). Battenfield et al. observed increasing accuracy in spring bread wheat with increasing training population size with the best model reaching 0.67 forward predictive accuracy (Battenfield et al., 2016). An accuracy range of 0.59-0.61 was observed when using an optimal model within a large wheat breeding program (Verges and Van Sanford, 2020). However, poor predictive accuracy for test weight was also shown in buckwheat (*Fagopyrum esculentum*) and wheat ranging from 0.26-0.36 and 0.2-0.3, respectively (Yabe et al., 2018; Borrenpohl et al., 2020). Even though GS can greatly reduce the number of generations or number of replications needed to develop varieties, it will decrease selection accuracy when compared to direct, phenotypic observation.

Recent advances in phenotyping proposes a new, more efficient method of prediction. Phenotyping is widely considered the bottleneck for genetic gain in plant breeding, whether breeders are making decisions directly based upon those phenotypes or creating prediction models from phenotypes (Bernardo, 2008; Rincent et al., 2012). As high-throughput phenotyping is becoming more practical, recent findings indicate that GS may be surpassed by a new, more efficient method: phenomic selection (PS) using near-infrared reflectance spectroscopy (NIRS). NIRS is prominently used throughout agricultural settings to measure grain composition of seeds and nutritional value of feed (Chen et al., 2013; Samadi et al., 2020). More importantly, it is regularly used in soybean breeding programs to collect phenotypes including protein, oil, starch, and sugar content. It would be a simple endeavor to replace GS with PS in soybean breeding programs by replacing genetic markers with NIR spectral data. PS would provide an even greater decrease in resources needed for varietal development than GS, but with improved accuracy and optimized methods.

Originally suggested by Rincent et al. in 2018, phenomic selection is founded on the idea that spectral reflectance also reflects the underlying genetic variance as well as the environmental influence on the phenotype (Rincent et al., 2018). This seminal research provided a proof on concept that measured NIR on wheat grain and tissue and poplar (*Populus nigra*) wood and demonstrated PS to be a low-cost and successful method for selection in breeding programs with a potential 81% increase in genetic gain. Several other prediction models have since been evaluated by Lane et al. for predicting maize yield from NIR on kernels with accuracies above 0.72, but they did not compare directly to GS models (Lane et al., 2020). In-season prediction of yield in soybean has

also shown promise by using machine learning to develop models from hyperspectral wavelengths (Parmley et al., 2019). Sandhu et al. combined phenomic wavelength data with genomic markers to improve prediction accuracy for protein content and yield in wheat (Sandhu et al., 2021). While all of these studies collect NIR spectral data with different equipment, they all utilize thousands of wavelength data points. However, not all NIR equipment generate this size dataset. Therefore, the objective of our study was to compare genomic and phenomic prediction models for test weight of soybean seeds in order to accelerate its breeding selection.

Materials and Methods

Plant Materials

A total of 500 soybean accessions were selected for the USDA Soybean Germplasm Collection to comprehensively represent genetic variability in maturity groups (MG) IV and V (Qin et al., 2017). Among these, a panel of 136 accessions produced enough seed to accurately measure test weight. These accessions were grown in 3 m two-row plots with 76 cm row spacing in Blacksburg, VA and 3 m four-row plots with 76 cm row spacing in Warsaw, VA in 2019 and 2020. Plots were organized based upon maturity and grown in a randomized complete block design with two blocks at each location. Each block began with two commercial checks, AG4403 and Ellis. Due to limited seed quantity in 2019, block replicates were merged within locations prior to measurement.

Data Collection

Test Weight Data

All seed samples were cleaned by removing moldy, mottled, discolored, split, or off-types seeds as well as any soil or debris. Test weight and moisture content was measured using the GAC 2500-AGRI Grain Analysis Computer (DICKEY-john, Auburn, IL). Test weight was reported as kilograms per hector liter (kg hL^{-1}) after being adjusted to a 13% moisture content (Liu et al., 2019). A minimum of 13 oz, or roughly 370 g, of soybean seed were used for accurate measurements. Three technical replicates were performed and averaged for each sample to account for equipment variation. Two accessions were removed from further analysis after abnormal test weight was observed.

Near-infrared Reflectance Spectroscopy Data

All samples were non-destructively analyzed using the DA 7250 NIR Analyzer spectrophotometer (PerkinElmer Inc.). Samples of 2.5 oz, or roughly 75 g, were placed in a small, rotating cup intended for use with whole seeds and scanned twice with a repack step between scans. The spectral dataset consisted of 141 wavelength data points collected between the 950-1650 nm range at 5 nm intervals. Average spectral absorbance values were exported and then treated using the Unscrambler X software (CAMO Analytics). The spectral wavelengths for each sample can be viewed in Figure 1. Similar to Rincent et al. and Lane et al., individual reflectance points were normalized by centered by the mean and scaled by the standard deviation and then the first derivative using the Savitky-Golay method (Savitzky and Golay, 1964; Rincent et al., 2018; Lane et al., 2020). A total of five accessions were removed from further analysis after inconsistent reflectance data was observed.

Genomic Data

Publicly available SNP marker data (www.soybase.org) of the 129 accessions were downloaded for the SoySNP50K SNPs data repository (Song et al., 2015). A total of 42,509 initial SNPs were filtered by low minor allele frequency (MAF <0.05) and missing genotypes, which resulted in 35,454 SNPs being used for analysis.

Evaluation of Genomic and Phenomic Selection

Genomic Prediction

GS was performed using gBLUP (genomic best linear unbiased prediction with the TASSEL 5 genomic selection function (Bradbury et al., 2007). A mixed model was used to incorporate a kinship matrix (K) and population structure (Q) for increased statistical power through the Q+K approach (Yu et al., 2006). Q was evaluated with a

discriminant analysis of principal components (DAPC) using the *adegenet* package in R to identify clusters of genetically related individuals (Jombart, 2008; Jombart et al., 2010). Successive k-means clustering was performed with the *find.clusters* function with maximum cluster set to $k = 40$. All principal components were retained, and Bayesian information criterion (BIC) was used to identify an optimal number of clusters. Then using the *dapc* function, an optimal number of principal components were retained to maximize cumulative variance without overfitting, and all discriminant functions were retained. A kinship matrix was create using the Centered_IBS method in TASSEL 5 (Bradbury et al., 2007; Endelman and Jannink, 2012). K-fold cross validation was performed using k = 5 with 10 iterations with r^2 collected for each environment. Environments with two blocks had 100 total iterations performed, 50 for each block.

Phenomic Prediction

JMP Pro 15.0.0 (SAS Institute Inc.) was used to perform statistical analyzes and predictions. Using an ANOVA with a 99% confidence interval ($P < 0.01$), location, year, accession, and the location × year (henceforth referred to as an environment) interaction variable were treated as fixed effects and a connecting letter report for accession and environment was generated post hoc using Tukey's honest significant difference (HSD). A T-test was used to generate connecting letter reports for location and year. Since our number of observations exceeded our phenomic predictors ($p < n$), a standard least squares regression was deemed suitable for prediction models. For similar reasons, PS accuracy was not estimated within locations as the decrease in total observations limited the least squares approach. PS accuracy within locations is of little interest as phenomic markers theoretically incorporate both genomic and environmental effects and should be

used to predict across environments. A total of 50 cross validations were performed using 80% training populations and 20% validation populations which were randomly assigned through JMP. The Pearson's correlation coefficient (r^2) was collected for each cross validation by comparing predicted and observed test weights for each validation population. Three cross validations were also performed to evaluate the predictive accuracy across locations and years; Blacksburg, VA observations, Warsaw, VA observations, and 2020 observations were used as training populations to predict test weight in Warsaw, VA, Blacksburg, VA, and 2019, respectively. Tukey's honest significant difference test was used to compare r^2 values between all average genomic and phenomic predictions to evaluate selection potential.

Results

Test Weight Observations

Test weight measurements exhibited normal, continuous distributions across all environments and averaged 69.5 kg hL $^{-1}$ with a standard deviation of 1.85 kg hL $^{-1}$. All samples were within 62.9 and 75.8 kg hL $^{-1}$. It should be noted that the standard U.S. test weight used for calculating soybean value is 74.8 kg hL $^{-1}$ (60 lbs per bushel), which suggests the potential for test weight improvement using accessions in our collection with high test weight as parental lines. Figure 2 displays frequency distributions for all environments (2a), 2019 Blacksburg, VA (2b), 2019 Warsaw, VA (2c), 2020 Blacksburg, VA (2d), and 2020 Warsaw, VA (2e) with means and standard deviations of 70.6, 71.2, 69.7, and 67.9 kg hL $^{-1}$ and 1.43, 1.31, 1.25, and 1.45 kg hL $^{-1}$, respectively. Accession, location, year, and environment were all shown to be significant effects for test weight ($P < 0.01$). Most accessions had statically similar means, but PI87059 and PI157487B (both originating from South Korea) displayed the highest and lowest average respective test weights, 72.9 and 67.3 kg hL $^{-1}$. When considering location and year as independent effects, Blacksburg, VA and 2019 displayed the highest means, 70.2 and 70.9 kg hL $^{-1}$ respectively. However, when considering each location and year combination as an environment, 2019 Warsaw, VA displayed the highest mean test weight (72.2 kg hL $^{-1}$) followed by 2019 Blacksburg, VA (70.6 kg hL $^{-1}$), 2020 Blacksburg, VA (69.8 kg hL $^{-1}$), and 2020 Warsaw, VA (67.9 kg hL $^{-1}$).

Genomic and Phenomic Prediction

Through DAPC, 80 principal components that accounted for 81.2% of cumulative variance were retained. Using BIC, $k = 3$ was determined as the optimal number of

clusters (Figure 3). Country of origin and MG for accessions within each cluster are shown in Table 1. Cluster I (n = 21) contained 20 MG IV accessions from the U.S. (95%) and one accession from South Korea (1%). Cluster II (n = 46) contained 41 accessions from China (89%) and a single accession each from Georgia, Indonesia, Japan, South Korea, and the U.S. (2.2% each). All but one accession in Cluster II belong to MG IV. Cluster III (n = 62) contained 32 accessions from China (51.6%), eight, five, six, and five accessions from Japan (12.9%), North Korea (8.1%), South Korea (9.7%), and the U.S (8.1%), respectively, and a single accession from Georgia, Nepal, Taiwan, Uganda, and Vietnam (1.6% each). Within Cluster III, 46 accessions belong to MG IV (75%) and 16 belong to MG V (25%).

gBLUP utilized the 35,454 SNPs and incorporated population structure to determine genomic estimated breeding values (GEBVs) which were compared to the observed test weight values. Overall, GP accuracy when using the 35,454 SNPs collection was poor, especially when predicting for average test weight across locations (r^2 = -0.04). Average predictive accuracy for individual environments were 0.26, 0.31, 0.07, and 0.17 for 2019 Blacksburg, VA, 2019 Warsaw, VA, 2020 Blacksburg, VA, and 2020 Warsaw, VA, respectively (Figure 4). Both environments from 2019 had statistically greater accuracies than all other genomic models, but they still exhibited much lower accuracy than the phenomic prediction model.

Cross validation (80% training and 20% validation population) over 50 iterations for phenomic prediction displayed an average predictive accuracy of 0.75 and an average root mean square error (RMSE) of 0.91. As evidenced in the boxplots of Figure 4, predictions based on NIRS data outperformed genomic models and generated precise

accuracy measurements across all iterations. Phenomic prediction accuracy was determined by generating predictions for the validation population from a model created with the training population and then comparing the predicted test weight value to the observed test weight value. Figure 5 shows four predicted test weight and observed test weight linear regression comparisons for cross validation iterations one through four as an example of the 50 iterations. Additionally, cross validation performed across locations and years showed variable accuracy. In Figure 6, correlations between predicted and observed test weights using models created from Blacksburg, VA to predict Warsaw, VA test weights (6a), Warsaw, VA to predict Blacksburg, VA test weights (6b), and 2020 to predict 2019 test weights (6c) are shown. The respective r^2 and RMSE values for each model were 0.59, 0.37, and 0.16 and 0.96, 1.64, 1.90.

Discussion

Test weight in soybean has historically been overlooked as a breeding objective and considered an afterthought until determining sales price. However, foundational research in other crops and recent investigations in soybean have identified test weight as an important factor for seed quality, processing, and transportation efficiency. As soybean breeders move toward developing varieties with improved test weight, they will face difficulties in phenotyping, especially during early breeding generations. To accurately and precisely measure test weight, a sufficient quantity of seed is required, and this quantity is unattainable until later breeding generations. This limits a breeder's ability to make selections for test weight in pre-breeding and early breeding materials. With this study, we evaluated the potential for GS and PS of test weight in soybean and identified that PS would be an ideal tool for achieving genetic gain with small seed quantity to measure test weight in early generations.

We observed consistent test weight performance, ranging from 62 to 76 kg hL^{-1}. However, we also observed significant environmental effect that is consistent with previous research (Liu et al., 2019; McNeece et al., 2021). This implies that improvement for soybean test weight will be influenced by G x E interactions, ultimately impacting breeder selections. Jarquin et al. suggested GS that utilizes high-density molecular marker sets, such as the marker set used in this project, combined with replicated, multi-location trials would be a suitable tool for making selections (Jarquin et al., 2016). While some researchers have reported GS accuracies for wheat test weight above 0.6 (Battenfield et al., 2016; Verges and Van Sanford, 2020), our results exhibit poor predictive ability for soybean test weight. Even when predicting within a single

environment, our strongest model only displays an average accuracy of 0.31. This aligns with similar studies in rice, wheat, and buckwheat (Isidro et al., 2015; Yabe et al., 2018; Borrenpohl et al., 2020). Once averaged across environments, that accuracy deteriorates to have no predictive ability (-0.04). Negative GS predictive ability has also been observed during a selection cycle for buckwheat test weight (Yabe et al., 2018). Through a combination of poor accuracy and environmental effects, GS appears to be a less-than-ideal method for accelerating genetic gain in soybean test weight.

Alternatively, PS is a recently published method that utilizes NIRS data to generate predictions for selection decisions. To beneficially impact genetic gains, PS will need to either decrease resources needed to make selections or showcase improved prediction ability when compared to GS. Rincent et al. identified a potential genetic gain increase of 81% when using PS over GS for grain yield in wheat (Rincent et al., 2018). This increase was a function of prediction accuracy as well as decreased cost. Lane et al. determined that PS would be economically suitable when NIR wavelengths can be easily obtained from each sample-environment combination (Lane et al., 2020). In soybean breeding programs, NIRS can be quickly obtained for each sample (often at no additional cost), as NIR analyzers are prominently used to measure seed composition such as protein and oil content.

Rincent et al. and Lane et al. reported variable predictive accuracies that depended on trait and environment, but general predictive abilities when including grain NIRS measurements from the same environment in both training and validation populations ranged from 0.3-0.6 for wheat yield and 0.72-0.84 for maize yield, respectively (Rincent et al., 2018; Lane et al., 2020). Rincent et al. compared accuracies directly to GS models,

whereas Lane et al. evaluated several PS models. It should be noted that these two studies utilized large wavelengths ranges with many repeated scans averaged together. Contrarily, by following the suggested method for collecting seed composition measurements using the DA 7250 NIR Analyzer (PerkenElmer, Inc.), our analysis only obtained 141 wavelength data points that were averaged over two scans. While this decreased our number of observations, it allowed us to use a standard least squares approach, simplifying the model design.

Our results exhibited similar success for PS potential, especially when compared to GS accuracy. Using a model that incorporated all environments, we observed an average predictive ability of 0.75, and the variability across cross validation iterations was markedly smaller than observed in GS (Figure 4). Thus, a PS model that incorporates NIRS data from all environments of interest could be an effective method for selection. As to be expected, however, the accuracy decreased when using separate environments as training and validation populations. When using Blacksburg, VA and Warsaw, VA to predict test weight for the other location, accuracy fell to 0.59 and 0.37, respectively. Similar accuracy drops between environments have been observed in wheat and maize (Rincent et al., 2018; Lane et al., 2020). The Blacksburg, VA dataset likely maintained a higher accuracy due to more consistent test weights between years. Another decline in accuracy (to 0.16) occurred when using 2020 spectra to predict for 2019. As 2019 had significantly higher test weights than 2020, the model underestimated the test weight for several accessions. Intriguingly, it also overestimated some test weights that were observed to be near average. Both locations in 2020 experienced greater precipitation rates than during 2019, and portions of seed filling in 2019 suffered drought. Excess

rainfall and soil water have been shown to negatively affect test weight in wheat (Guarienti et al., 2005), and moisture content during seed development and maturation directly affects physical properties of soybean seed (Tunde-Akintunde et al., 2005). This may suggest limited phenomic predictive accuracy when certain environmental variables are inconsistent. However, when compared to the lack of GS predictive ability, all PS models still provide utility when averaging test weight across locations.

Overall, our results highlight the potential benefits of incorporating phenomic prediction models for breeders to make selections for soybean test weight and exhibit increased accuracy of PS when compared to GS. In one swift motion, breeders can use a common NIR analyzer to measure seed composition and collect NIRS spectra to obtain high predictive accuracy across multi-location trials. These results also support the growing body of research highlighting PS as a useful tool for plant breeders working in many crops and selecting for several traits.

References

Battenfield, S.D., C. Guzmán, R.C. Gaynor, R.P. Singh, R.J. Peña, et al. 2016. Genomic Selection for Processing and End-Use Quality Traits in the CIMMYT Spring Bread Wheat Breeding Program. The Plant Genome 9(2): plantgenome2016.01.0005. doi: 10.3835/plantgenome2016.01.0005.

Bernardo, R. 2008. Molecular Markers and Selection for Complex Traits in Plants: Learning from the Last 20 Years. Crop Science 48(5): 1649–1664. doi: 10.2135/cropsci2008.03.0131.

Borrenpohl, D., M. Huang, E. Olson, and C. Sneller. 2020. The value of early-stage phenotyping for wheat breeding in the age of genomic selection. Theor Appl Genet 133(8): 2499–2520. doi: 10.1007/s00122-020-03613-0.

Bradbury, P.J., Z. Zhang, D.E. Kroon, T.M. Casstevens, Y. Ramdoss, et al. 2007. TASSEL: software for association mapping of complex traits in diverse samples. Bioinformatics 23(19): 2633–2635. doi: 10.1093/bioinformatics/btm308.

Chen, J., X. Ren, Q. Zhang, X. Diao, and Q. Shen. 2013. Determination of protein, total carbohydrates and crude fat contents of foxtail millet using effective wavelengths in NIR spectroscopy. Journal of Cereal Science 58(2): 241–247. doi: 10.1016/j.jcs.2013.07.002.

Crossa, J., P. Pérez-Rodríguez, J. Cuevas, O. Montesinos-López, D. Jarquín, et al. 2017. Genomic Selection in Plant Breeding: Methods, Models, and Perspectives. Trends in Plant Science 22(11): 961–975. doi: 10.1016/j.tplants.2017.08.011.

Ding, J.Q., J.L. Ma, C.R. Zhang, H.F. Dong, Z.Y. Xi, et al. 2011. QTL mapping for test weight by using F2:3 population in maize. J Genet 90(1): 75–80. doi: 10.1007/s12041-011-0036-3.

Endelman, J.B., and J.-L. Jannink. 2012. Shrinkage Estimation of the Realized Relationship Matrix. G3 (Bethesda) 2(11): 1405–1413. doi: 10.1534/g3.112.004259.

Ghaderi, A., E.H. Everson, and W.T. Yamazaki. 1971. Test Weight in Relation to the Physical and Quality Characteristics of Soft Winter Wheat (Triticum aestivum L. em Thell)1. Crop Science 11(4): cropsci1971.0011183X001100040015x. doi: 10.2135/cropsci1971.0011183X001100040015x.

Markets Insider. Soybeans PRICE Today. https://markets.businessinsider.com/commodities/soybeans-price (accessed 21 September 2021).

Guarienti, E.M., C.F. Ciacco, G.R. da Cunha, L. de J.A. Del Duca, and C.M. de O. Camargo. 2005. Effects of rainfall, relative air humidity and excess and soil water deficit on hectoliter weight, thousand grain weight and wheat grain yield. Food Sci. Technol 25: 412–418. doi: 10.1590/S0101-20612005000300004.

Isidro, J., J.-L. Jannink, D. Akdemir, J. Poland, N. Heslot, et al. 2015. Training set optimization under population structure in genomic selection. Theoretical and Applied Genetics 1(128): 145–158. doi: 10.1007/s00122-014-2418-4.

Jarquin, D., J. Specht, and A. Lorenz. 2016. Prospects of Genomic Prediction in the USDA Soybean Germplasm Collection: Historical Data Creates Robust Models for Enhancing Selection of Accessions. G3: Genes, Genomes, Genetics 6(8): 2329–2341. doi: 10.1534/g3.116.031443.

Jing, Q., D. Jiang, T. Dai, and W. Cao. 2003. [Effects of genotype and environment on wheat grain quality and protein components]. Ying Yong Sheng Tai Xue Bao 14(10): 1649–1653.

Jombart, T. 2008. adegenet: a R package for the multivariate analysis of genetic markers. Bioinformatics 24(11): 1403–1405. doi: 10.1093/bioinformatics/btn129.

Jombart, T., S. Devillard, and F. Balloux. 2010. Discriminant analysis of principal components: a new method for the analysis of genetically structured populations. BMC Genetics 11(1): 94. doi: 10.1186/1471-2156-11-94.

Kaya, Y., and M. Akcura. 2014. Effects of genotype and environment on grain yield and quality traits in bread wheat (T. aestivum L.). Food Sci. Technol 34: 386–393. doi: 10.1590/fst.2014.0041.

Klein, S.J., M.A. Smith, and K.J. Frey. 1993. Recurrent Selection for Test Weight and Grain Yield of Oat. Crop Science 33(4): cropsci1993.0011183X003300040022x. doi: 10.2135/cropsci1993.0011183X003300040022x.

Lane, H.M., S.C. Murray, O.A. Montesinos-López, A. Montesinos-López, J. Crossa, et al. 2020. Phenomic selection and prediction of maize grain yield from near-infrared reflectance spectroscopy of kernels. The Plant Phenome Journal 3(1): e20002. doi: 10.1002/ppj2.20002.

Liu, L., E. Prenger, J. Zhang, B. Little, M.A.R. Mian, et al. 2019. Impact of genotype, seed composition, agronomic trait, and environment on soybean test weight. Journal of Crop Improvement 33(6): 711–729. doi: 10.1080/15427528.2019.1659205.

McNeece, B.T., J.H. Gillenwater, Z. Li, and M.A.R. Mian. 2021. Assessment of soybean test weight among genotypes, environments, agronomic and seed compositional traits. Agronomy Journal 113(3): 2303–2313. doi: 10.1002/agj2.20665.

Mississippi Soybean Promotion Board. Soybean Grade Requirements and Discount Schedules. https://www.mssoy.org/wp-content/uploads/2015/09/SOYBEAN-DISCOUNT-SCHEDULES.pdf.

Naeve, S.L., and J. Miller-Garvin. 2019. United States Soybean Quality. American Soybean Association.

Parmley, K., K. Nagasubramanian, S. Sarkar, B. Ganapathysubramanian, and A.K. Singh. 2019. Development of Optimized Phenomic Predictors for Efficient Plant Breeding Decisions Using Phenomic-Assisted Selection in Soybean. Plant Phenomics 2019. doi: 10.34133/2019/5809404.

Pixley, K.V., and K.J. Frey. 1991. Inheritance of Test Weight and Its Relationship with Grain Yield of Oat. Crop Science 31(1): cropsci1991.0011183X003100010008x. doi: 10.2135/cropsci1991.0011183X003100010008x.

Qin, J., Q. Song, A. Shi, S. Li, M. Zhang, et al. 2017. Genome-wide association mapping of resistance to Phytophthora sojae in a soybean [Glycine max (L.) Merr.] germplasm panel from maturity groups IV and V. PLOS ONE 12(9): e0184613. doi: 10.1371/journal.pone.0184613.

Rharrabti, Y., S. Elhani, V. Martos-Núñez, and L.F. García del Moral. 2001. Protein and Lysine Content, Grain Yield, and Other Technological Traits in Durum Wheat under Mediterranean Conditions. J. Agric. Food Chem. 49(8): 3802–3807. doi: 10.1021/jf001139w.

Rincent, R., J.-P. Charpentier, P. Faivre-Rampant, E. Paux, J.L. Gouis, et al. 2018. Phenomic Selection Is a Low-Cost and High-Throughput Method Based on Indirect Predictions:

Proof of Concept on Wheat and Poplar. G3: Genes, Genomes, Genetics 8(12): 3961–3972. doi: 10.1534/g3.118.200760.

Rincent, R., D. Laloë, S. Nicolas, T. Altmann, D. Brunel, et al. 2012. Maximizing the Reliability of Genomic Selection by Optimizing the Calibration Set of Reference Individuals: Comparison of Methods in Two Diverse Groups of Maize Inbreds (Zea mays L.). Genetics 192(2): 715–728. doi: 10.1534/genetics.112.141473.

Samadi, S. Wajizah, and A.A. Munawar. 2020. Near infrared spectroscopy (NIRS) data analysis for a rapid and simultaneous prediction of feed nutritive parameters. Data in Brief 29: 105211. doi: 10.1016/j.dib.2020.105211.

Sandhu, K.S., P.D. Mihalyov, M.J. Lewien, M.O. Pumphrey, and A.H. Carter. 2021. Combining Genomic and Phenomic Information for Predicting Grain Protein Content and Grain Yield in Spring Wheat. Frontiers in Plant Science 12: 170. doi: 10.3389/fpls.2021.613300.

Savitzky, Abraham., and M.J.E. Golay. 1964. Smoothing and Differentiation of Data by Simplified Least Squares Procedures. Anal. Chem. 36(8): 1627–1639. doi: 10.1021/ac60214a047.

Song, Q., D.L. Hyten, G. Jia, C.V. Quigley, E.W. Fickus, et al. 2015. Fingerprinting Soybean Germplasm and Its Utility in Genomic Research. G3: Genes, Genomes, Genetics 5(10): 1999–2006. doi: 10.1534/g3.115.019000.

Soy Transportation Coalition. Cargo Capacity. https://www.soytransportation.org/Stats/Modal_CargoCapacity.pdf (accessed 14 September 2021).

Teich, A.H. 1984. Heritability of grain yield, plant height and test weight of a population of winter wheat adapted to Southwestern Ontario. Theor Appl Genet 68(1–2): 21–23. doi: 10.1007/BF00252304.

The American Soybean Association. 2020. U.S. Exports: Soy Product Exports by Customer. http://soystats.com/u-s-exports-soy-products-exports-by-customersda/ (accessed 14 September 2021).

Tunde-Akintunde, T.Y., J.O. Olajide, and B.O. Akintunde. 2005. Mass-Volume-Area Related and Mechanical Properties of Soybean as a Function of Moisture and Variety. International Journal of Food Properties 8(3): 449–456. doi: 10.1080/10942910500267513.

U.S. Soybean Export Council. 2015. U.S. Soy: International Buyer's Guide. American Soybean Association-International Marketing (ASA-IM), St. Louis, MO.

Verges, V.L., and D.A. Van Sanford. 2020. Genomic Selection at Preliminary Yield Trial Stage: Training Population Design to Predict Untested Lines. Agronomy 10(1): 60. doi: 10.3390/agronomy10010060.

Wilson, R.F. 2004. Seed Composition. Soybeans: Improvement, Production, and Uses. 3rd ed. ASA, CSSA, SSSA, Madison, WI. p. 621–669

Yabe, S., T. Hara, M. Ueno, H. Enoki, T. Kimura, et al. 2018. Potential of Genomic Selection in Mass Selection Breeding of an Allogamous Crop: An Empirical Study to Increase Yield of Common Buckwheat. Frontiers in Plant Science 9: 276. doi: 10.3389/fpls.2018.00276.

Yu, J., G. Pressoir, W.H. Briggs, I. Vroh Bi, M. Yamasaki, et al. 2006. A unified mixed-model method for association mapping that accounts for multiple levels of relatedness. Nat Genet 38(2): 203–208. doi: 10.1038/ng1702.

Table 1. Countries of origin and maturity groups (MG) for clustered accessions as determined by discriminant analysis of principal components (DAPC).

	Cluster I (n = 21)		Cluster II (n = 46)		Cluster III (n = 62)		
	Count	%	Count	%	Count	%	Total
China	-	-	41	89	32	51.6	73
Georgia	-	-	1	2.2	1	1.6	2
Indonesia	-	-	1	2.2	-	-	1
Japan	-	-	1	2.2	8	12.9	9
Nepal	-	-	-	-	1	1.6	1
North Korea	-	-	-	-	5	8.1	5
South Korea	1	5	1	2.2	6	9.7	8
Taiwan	-	-	-	-	1	1.6	1
Uganda	-	-	-	-	1	1.6	1
United States	20	95	1	2.2	5	8.1	26
Vietnam	-	-	-	-	1	1.6	1
Unknown	-	-	-	-	1	1.6	1
MG IV	21	100	45	97.8	46	75	112
MG V	-	-	1	2.2	16	25	17

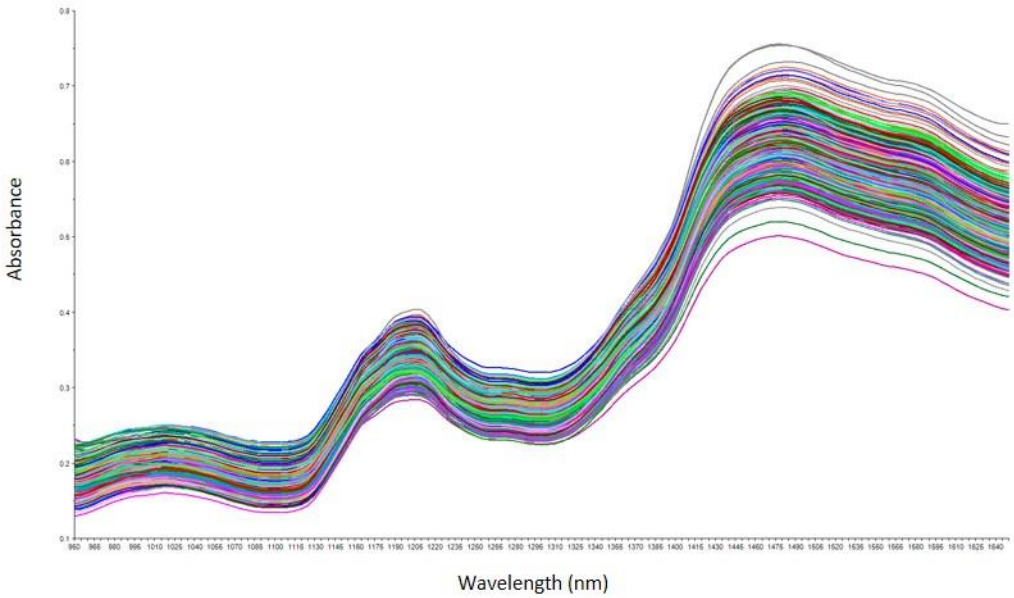

Figure 1. All untreated analyzed spectra from each sample of whole soybean seeds. Each colored line represents a unique accession-environment combination.

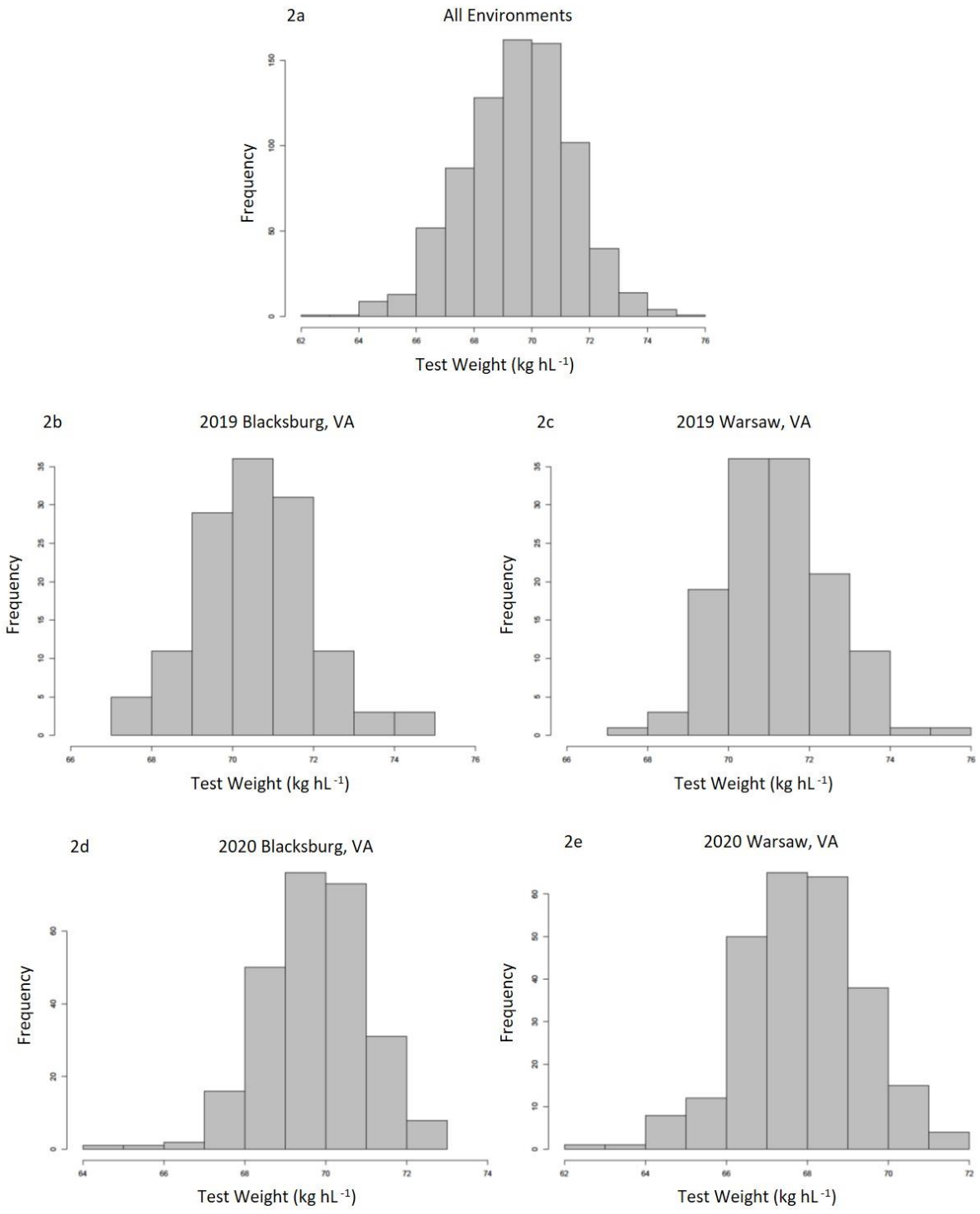

Figure 2. Frequency distributions displaying test weight measurements collected from all environments (2a), 2019 Blacksburg, VA (2b), 2019 Warsaw, VA (2c), 2020 Blacksburg, VA (2d), and 2020 Warsaw, VA (2e).

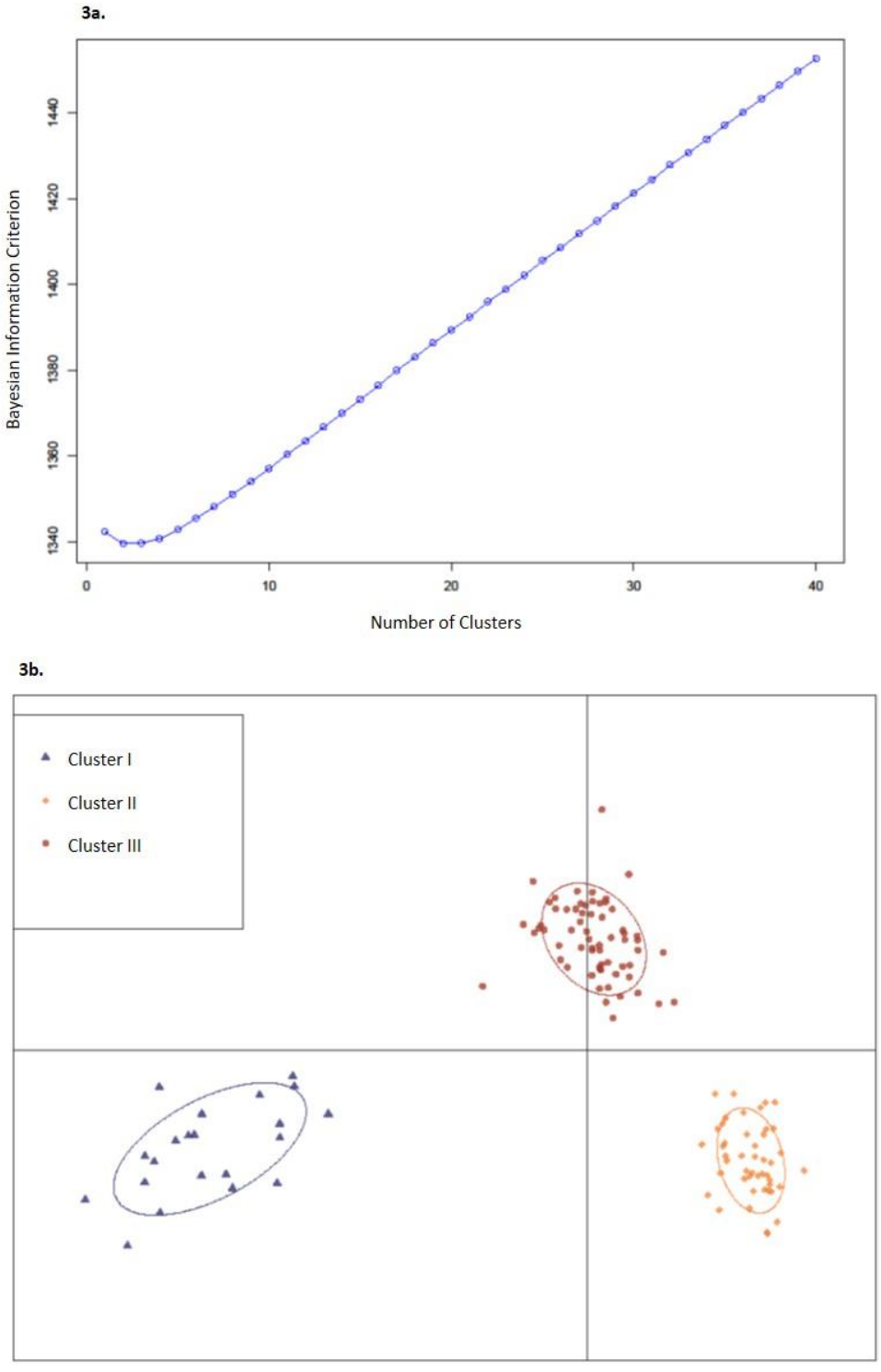

Figure 3. (3a) Bayesian information criterion for selecting optimal number of clusters. (3b) A scatter plot depicting the 3 clusters (k = 3) identified as likely subpopulations within the 129 accessions: cluster I (blue triangles, n = 21), cluster II (gold diamonds, n = 46), cluster III (red circles, n = 62).

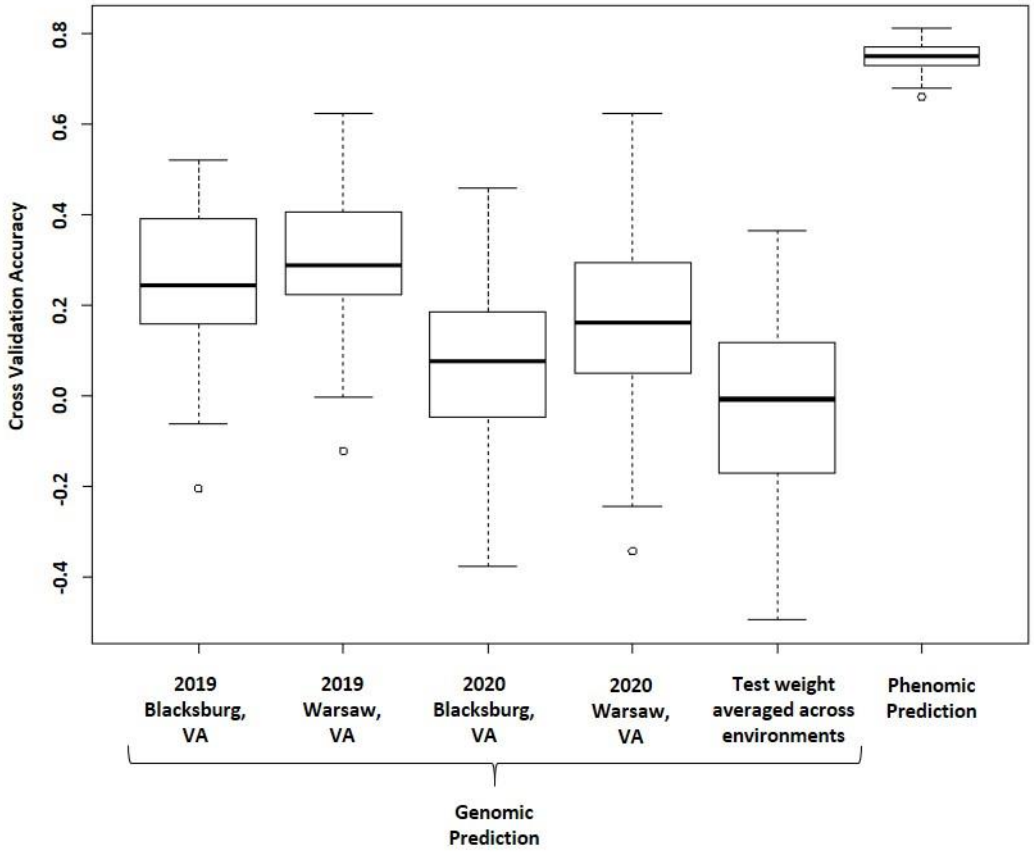

Figure 4. All cross validation iterations for each genomic prediction model used across environments and phenomic prediction using near-infrared reflectance spectroscopy. The outer lines of each box represent the upper and lower quartiles, and the bold line inside each box displays the median for each model. White circles represent outliers that fell outside the interquartile range.

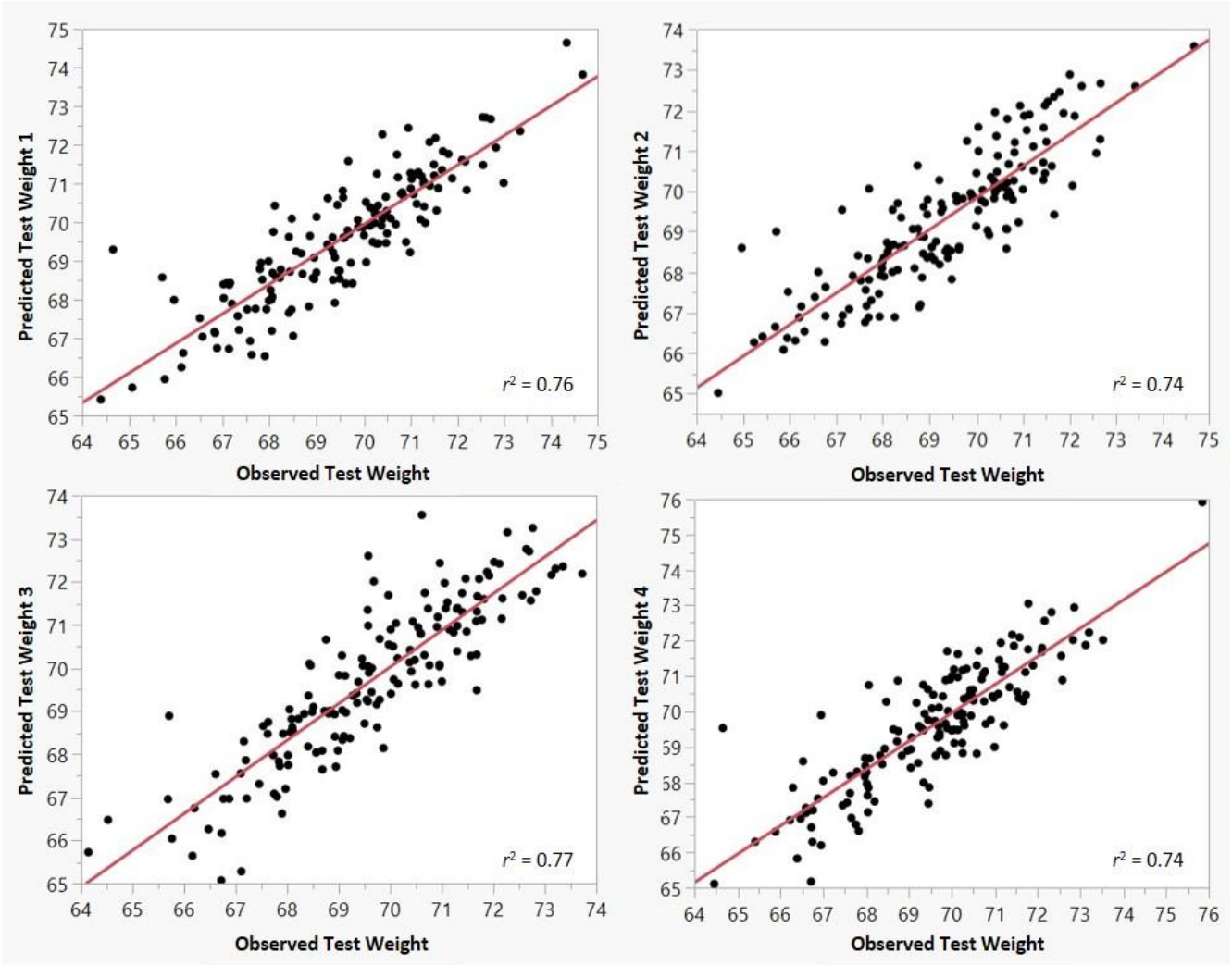

Figure 5. The correlations between observed test weight and predicted test weight for the first four phenomic prediction cross validation iterations.

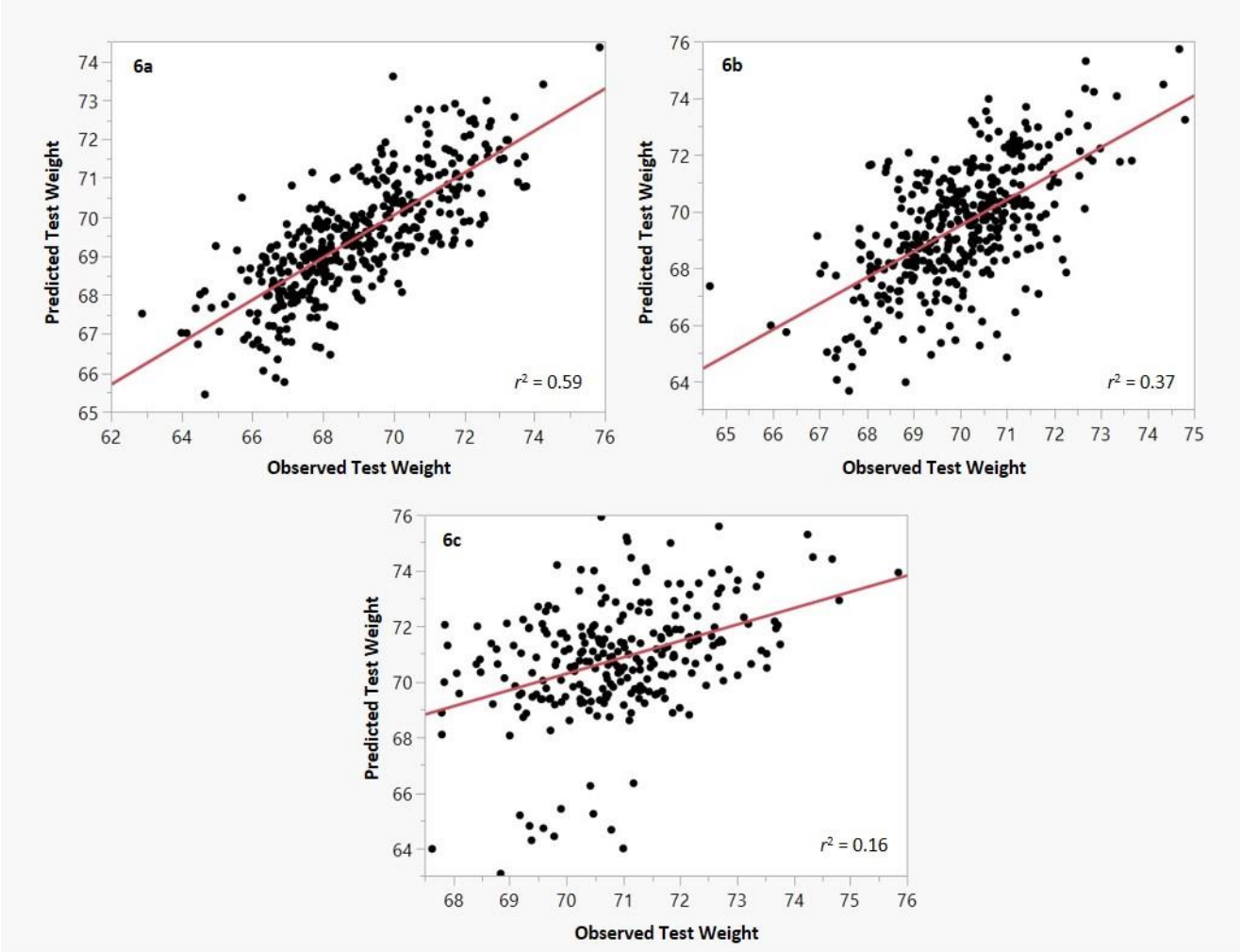

Figure 6. The correlations between observed test weight and predicted test weight by phenomic prediction cross validations across year and location. Models using Blacksburg, VA data to predict Warsaw, VA test weight (6a), Warsaw, VA data to predict Blacksburg, VA test weight (6b), and 2020 data to predict 2019 test weights (6c) are shown.

Chapter 4: Soybean Amino Acid and Fatty Acid Response to Sulfur Fertilization

William M. Singer[1], Keren Brooks[1], Mark Reiter[2], David L. Holshouser, and Bo Zhang[1]

[1]School of Plant and Environmental Sciences, Virginia Tech

[2] Virginia Tech Eastern Shore Agricultural Research and Extension Center, Virginia Tech

[3]Virginia Tech Tidewater Agricultural Research and Extension Center, Virginia Tech

Abstract

Soybean (*Glycine max*) is a valuable crop with end-use versatility in feed, food, and industrial purposes due to its amino acid and fatty acid profiles. With sulfur availability becoming increasingly important and the value of soybean intrinsically tied to these seed components, the objective of this project was to identify the impact of sulfur fertilization on the monomers of seed protein and oil. A replicated field study was implemented in 2019 and 2020 in six locations across the state of Virginia using four forms of sulfur fertilizer: ammonium sulfate (AMS), ammonium thiosulfate (ATS), calcium sulfate (CS), and elemental sulfur (ES). Amino and fatty acids were quantified using near-infrared reflectance spectroscopy. Location proved to be a significant effect ($P < 0.001$) for all amino and fatty acids observed, while year influenced nine of the traits ($P < 0.001$). Sulfur fertilization only statically affected methionine concentrations ($P < 0.05$) with AMS, the most available source of sulfate, exhibiting the highest mean (0.565% methionine in seed, dry weight basis). Interaction between source and location variables suggests that site-specific implications may also be relevant.

Introduction

Soybean (*Glycine max*) is a valuable and versatile global crop due to its seed composition (Wilson, 2004). Protein and oil content as well as their respective monomers, amino acids and fatty acids, provide a variety of end-uses after processing. Soybean is regularly used in livestock feed as a protein source (Buttery and D'Mello, 1994; Boisen et al., 2000) because of the presence of all nine essential amino acids (Kuiken and Lyman, 1949; Goldflus et al., 2006). Many human food and industrial products also result from soybean oil (Liu, 1997), especially in high oleic acid varieties (Zambelli, 2021). Sulfur is a known nutrient required for optimal crop growth (Hitsuda et al., 2005; Franzen and Grant, 2008) and has been shown to impact soybean production (Hitsuda et al., 2008). Alongside nitrogen, sulfur is critical for seed protein formation and therefore influences seed component accumulation.

While sulfur deficiencies are typically not a concern, recent sulfur deposition data posits that increased sulfur fertilization may be needed in the future (USEPA, 2020). Kaiser and Kim (2013), Letham et al. (2021), and Cannon et al. (2021) identified limited to no agronomic improvement from sulfur application to soybean. Regardless, many physiological processes are dependent upon sulfur availability. Dev and Saggar (1974) recognized nitrogen to sulfur ratios were affected by sulfur fertilization and varied between soybean varieties. Additionally, nitrogen to sulfur cotyledon ratios and storage protein abundance were reportedly impacted by sulfur availability which suggests direct implications for overall amino acid biosynthesis (Sexton et al., 1998). This necessitates a special consideration of sulfur-containing amino acids, i.e. methionine and cysteine.

Methionine is one of the nine essential amino acids, required for metabolic processes, and is typically known as the initiating amino acid in proteins sequences (Brosnan et al., 2007). Conde-Aguilera et al. (2013) and Wu et al. (2012) showed limited methionine in poultry diets negatively impacts body composition and disease immunity, while methionine is the limiting amino acid in soybean proteins (Berry et al., 1962). Cysteine, while not one of the nine essential amino acids, plays a critical role in tertiary protein structure through disulfide bonds (Jessop et al., 2004). Limited cysteine in poultry diets also interferes with keratin production and thereafter feather production and quality (Wylie et al., 2001). As 36.6 million short tons of soybean meal used for livestock feed in the United States in 2020 (The American Soybean Association, 2020), it is important to explore how to increase the content of sulfur-containing amino acids.

Sulfur fertilizer can be applied in several different forms including ammonium sulfate (AMS), ammonium thiosulfate (ATS), elemental sulfur (ES), and calcium sulphate (CS) which all eventually oxidize to form plant-available sulfate (SO_4^-). AMS is the most available form; it requires little to no oxidation but is highly susceptible to leaching (Riley et al., 2002). ES has the slowest oxidation rate; however, it may be applied for delayed sulfur availability and decreased cost (Grant et al., 2012; Goyal et al., 2021). Ham et al. (1975) and Devi et al. (2012) recognized impacts on seed composition with sulfur fertilization, but very little research determined the impact of sulfur fertilization on protein and oil monomers for a variety of sources.

The research objective of this project was to determine the impact of different sulfur fertilizers on amino and fatty acid concentrations in soybean seed. It was

hypothesized that sulfur containing amino acids would be specifically affected by fertilization.

Materials and Methods

Site Descriptions

Fields with sandy loam soils and potential sulfur deficiencies were chosen across eastern Virginia. All locations were rain-fed and a full-season crop except for Suffolk 2019 and Eastville 2020, which were double-crop after wheat. Pests were controlled on a site-by-site basis.

Experimental Design

The experiment was grown as a replicated complete block design (RCBD) in four locations, Chesapeake, Essex, Painter, and Suffolk, Virginia, in 2019 and 2020, as well as Eastville and Virginia Beach, Virginia in 2020. Soybean varieties were selected individually for each site. In 2019, plot sizes were 300 ft^2, while in 2020, plot size was increased to 400 ft^2, with the exception of Suffolk which had a plot size of 255 ft^2 both years. Most locations were planted with a 15-inch row spacing, except for Chesapeake and Painter which were planted with 30-inch row spacing both years. Granular sulfur was broadcast evenly by hand as three different sources: AMS, ES, and CS. ATS, a liquid sulfur source, was applied using a hand sprayer. All sources were applied at three different rates: 10, 20, and 30 S lb $acre^{-1}$. Each block also contained an untreated control (UTC) as well as three rates of urea (9, 18, and 26 N lb $acre^{-1}$). Urea served as a positive nitrogen control for AMS and ATS which contained 9, 18, and 26 and 10, 20, and 30 N lb $acre^{-1}$, respectively.

Data collection

Near-infrared spectroscopy (NIRS) was used to quantify seed components (Pazdernik et al., 1997; Kovalenko et al., 2006), including amino acids cysteine,

histidine, isoleucine, leucine, lysine, methionine, phenylalanine, threonine, tryptophan, and valine as well as fatty acids linolenic, oleic, palmitic, and stearic acid. The DA 7250 NIR Analyzer and 2019 calibrations were utilized from PerkinElmer to quantify on a % in seed, dry weight basis. A small rotating tray was filled with whole soybean seeds and then repacked as a technical replicate.

Statistical analysis

Data were fitted with a mixed model ANOVA using standard least squares in JMP Pro 15.0.0 (SAS Institute Inc. 2019). Fixed effects included growing location, year, sulfur source, and the sulfur source × location interaction variable at the 95% confidence level ($P < 0.05$). Sulfur rate was included in a preliminary model, however due to singularity errors, it was removed from the final model. Connecting letter reports for location and source were generated post hoc using Least Square Means through Tukey's honest significant difference (HSD). The Student's T-test was used for the year effect. The test slice function in JMP Pro 15.0.0 (SAS Institute Inc. 2019) was used to identify significant contributions to the sulfur source × location interaction variable.

Results

All traits exhibited a significant model when including all effects as shown in Table 1. However, linolenic acid exhibited a poor R-squared value, while oleic acid exhibited a poor root mean square error (RMSE). Location was a significant factor for all amino and fatty acid traits ($P < 0.001$). Year was a significant factor for nine traits ($P < 0.001$ or 0.01), excluding isoleucine, leucine, methionine, and valine. The sulfur source × location interaction variable was significant for cysteine, histidine, isoleucine, leucine, lysine, phenylalanine ($P < 0.05$), and threonine ($P < 0.01$). Sulfur source was only a significant factor for methionine ($P < 0.05$).

Significant differences between amino acid (Table 2) and fatty acid (Table 3) concentration means were observed for all traits. Chesapeake, VA displayed the highest amino acid concentrations for histidine, isoleucine, leucine, lysine, phenylalanine, and valine with means of 1.09, 2.01, 3.22, 2.72, 2.18, and 2.08% in seed, respectively. Threonine performance was identical in Chesapeake and Painter, VA ($\mu = 1.56\%$ in seed) which was significantly greater than all other locations. Cysteine and methionine concentrations were highest in Painter, VA with means of 0.58 and 0.572% in seed, respectively. However, Painter, VA cysteine levels were statistically indifferent from seeds harvested in Eastville, VA, whereas methionine levels were also indifferent from levels observed in Virginia Beach, VA. Tryptophan concentrations were highest in Eastville, VA ($\mu = 0.42\%$ in seed).

Performance for fatty acids was more similar across locations. Essex and Painter, VA linolenic acid levels ($\mu = 6.83$, 6.87% in seed, respectively) were significantly higher than measured values from Chesapeake, VA ($\mu = 6.55\%$ in seed) but statistically similar

to all other locations. Painter, VA displayed the highest oleic acid concentrations with a mean of 26.74% in seed, ≥ 2% more than all other locations. Eastville, Painter, and Virginia Beach, VA all displayed the greatest levels of palmitic acid with means of 11.36, 11.55, 11.41% in seed, respectively. Stearic acid concentrations were significantly higher in Virginia Beach, VA (μ = 3.99% in seed).

Amino and fatty acids concentrations were typically higher from seeds produced in 2019 as shown in Tables 4 and 5. Cysteine, histidine, threonine, and tryptophan as well as oleic, palmitic, and stearic acid all displayed higher means in 2019 (μ = 0.58, 1.08, 1.54, 0.41, 24.94, 11.32, and 3.72% in seed, respectively). Lysine, phenylalanine, and linolenic acid had higher concentrations in 2020 with means of 2.68, 2.12, and 6.87, respectively.

Sulfur source proved to only be significant for methionine in which AMS performed significantly higher than ES and urea with means of 0.565, 0.56, and 0.56% in seed, respectively (Table 6). CS (μ = 0.563% in seed), ATS (μ = 0.563% in seed), and UTC (μ = 0.562% in seed) were statistically indifferent from all other treatments. Figure 1 displays the normal distribution of methionine with a mean and median of 0.56% in seed and a standard deviation of 0.014. Table 7 exhibits that Chesapeake and Painter, VA and all sources were consistently, significant contributors to the interaction effect for amino acid concentrations. Excluding methionine, tryptophan, and valine, all source treatments were deemed significant interaction components ($P < 0.001$). Chesapeake, VA exhibited interaction significance for cysteine ($P < 0.05$), leucine ($P < 0.01$), lysine ($P < 0.05$), phenylalanine ($P < 0.05$), and threonine ($P < 0.001$). Painter, VA also exhibited interaction significance for leucine ($P < 0.05$), lysine ($P < 0.01$), and phenylalanine ($P <$

0.05) as well as histidine ($P < 0.01$) and isoleucine ($P < 0.01$). Additionally, Essex, VA displayed interaction importance for cysteine ($P < 0.01$).

Discussion

Amino and fatty acid concentrations were significantly affected by the growing location and year more consistently than sulfur fertilization source. Similar results were found in Goyal et al. (2021). Chesapeake and Painter, VA appeared to have improved performances when compared to other locations. This is likely due to soil factors or varietal differences. This may suggest site-specific implications for soybeans grown for particular end-uses such as high oleic acid, high meal protein, or high methionine varieties. Methionine was the sole trait impacted by sulfur source which aligns with our hypothesis. Surprisingly, cysteine did not show a response to sulfur fertilization. As cysteine is a precursor to methionine in plant amino acid biosynthesis (Bonner et al., 2005), an increase in cysteine from sulfate uptake may not be evident in post-harvest evaluation.

Plots treated with AMS had the highest concentration of methionine, possibly as a result of higher SO_4^- availability. It was statistically greater than ES, the slowest oxidizing source, and urea, the positive control for nitrogen. This suggests the increase in methionine resulted from the added sulfur and not nitrogen. AMS was not statistically different than ATS and CS, intermediate SO_4^- sources, but confoundingly, it was not significantly more than UTC. The availability of SO_4^- appears to be playing an important role, and combined with recent research elucidating season-long nutrient uptake and partitioning trends in soybean (Bender et al., 2015; Gaspar et al., 2018), further investigation is required to determine impacts of sulfur application timing. Moreover, dependence found within the sulfur source × location interaction variable further indicates site-specific implications.

Additionally, while higher levels of methionine were observed, the increased amount would not alter end-use quality on a basis of poultry feed requirements (Fernandez et al., 1994; Bunchasak, 2009). Given these results and the lack of significance for other amino and fatty acids, we are unable to suggest sulfur application of any source for seed composition motives. However, significant results in methionine suggest that future research could identify soybean development stages and SO_4^- availability timing to optimize sulfur-containing amino acid concentrations.

References

Bender, R.R., J.W. Haegele, and F.E. Below. 2015. Nutrient Uptake, Partitioning, and Remobilization in Modern Soybean Varieties. Agronomy Journal 107(2): 563–573. doi: 10.2134/agronj14.0435.

Berry, T.H., D.E. Becker, O.G. Rasmussen, A.H. Jensen, and H.W. Norton. 1962. The Limiting Amino Acids in Soybean Protein. Journal of Animal Science 21(3): 558–561. doi: 10.2527/jas1962.213558x.

Boisen, S., T. Hvelplund, and M.R. Weisbjerg. 2000. Ideal amino acid profiles as a basis for feed protein evaluation. Livestock Production Science 64(2): 239–251. doi: 10.1016/S0301-6226(99)00146-3.

Bonner, E.R., R.E. Cahoon, S.M. Knapke, and J.M. Jez. 2005. Molecular Basis of Cysteine Biosynthesis in Plants: STRUCTURAL AND FUNCTIONAL ANALYSIS OF O-ACETYLSERINE SULFHYDRYLASE FROM ARABIDOPSIS THALIANA *. Journal of Biological Chemistry 280(46): 38803–38813. doi: 10.1074/jbc.M505313200.

Brosnan, J.T., M.E. Brosnan, R.F.P. Bertolo, and J.A. Brunton. 2007. Methionine: A metabolically unique amino acid. Livestock Science 112(1): 2–7. doi: 10.1016/j.livsci.2007.07.005.

Bunchasak, C. 2009. Role of Dietary Methionine in Poultry Production. The Journal of Poultry Science 46(3): 169–179. doi: 10.2141/jpsa.46.169.

Buttery, P.J., and J.P.F. D'Mello. 1994. Amino Acid Metabolism in Farm Animals: An Overview. Amino Acids in Farm Animal Nutrition. CAB International, Wallingford, United Kingdom. p. 1–10

Cannon, K., M.A. McClure, X. Yin, and C. Sams. 2021. Corn and soybean response to sulfur fertilizer in West Tennessee. Crop, Forage & Turfgrass Management 7(1): e20092. doi: 10.1002/cft2.20092.

Conde-Aguilera, J.A., C. Cobo-Ortega, S. Tesseraud, M. Lessire, Y. Mercier, et al. 2013. Changes in body composition in broilers by a sulfur amino acid deficiency during growth. Poult. Sci. 92(5): 1266–1275. doi: 10.3382/ps.2012-02796.

Dev, G., and S. Saggar. 1974. Effect of Sulfur Fertilization on the N-S Ratio in Soybean Varieties. Agronomy Journal 66(3): 454–456. doi: 10.2134/agronj1974.00021962006600030034x.

Devi, K.N., L.N.K. Singh, M. Sumarjit Singh, S. Basanta Singh, and K. Khamba Singh. 2012. Influence of Sulphur and Boron Fertilization on Yield, Quality, Nutrient Uptake and Economics of Soybean (Glycine max) under Upland Conditions. JAS 4(4): p1. doi: 10.5539/jas.v4n4p1.

Fernandez, S.R., S. Aoyagi, Y. Han, C.M. Parsons, and D.H. Baker. 1994. Limiting order of amino acids in corn and soybean meal for growth of the chick. Poult. Sci. 73(12): 1887–1896. doi: 10.3382/ps.0731887.

Franzen, D., and C.A. Grant. 2008. Sulfur Response Based on Crop, Source, and Landscape Position. Sulfur: A Missing Link between Soils, Crops, and Nutrition. John Wiley & Sons, Ltd. p. 105-116–4

Gaspar, A.P., C.A.M. Laboski, S.L. Naeve, and S.P. Conley. 2018. Secondary and Micronutrient Uptake, Partitioning, and Removal across a Wide Range of Soybean Seed Yield Levels. Agronomy Journal 110(4): 1328–1338. doi: 10.2134/agronj2017.12.0699.

Goldflus, F., M. Ceccantini, and W. Santos. 2006. Amino acid content of soybean samples collected in different Brazilian states. Brazilian Journal of Poultry Science 8(2): 105–111. doi: 10.1590/S1516-635X2006000200006.

Goyal, D., D.W. Franzen, and A. Chatterjee. 2021. Do crops' responses to sulfur vary with its forms? Agrosystems, Geosciences & Environment 4(3): e20201. doi: 10.1002/agg2.20201.

Grant, C.A., S.S. Mahli, and R.E. Karamanos. 2012. Sulfur management for rapeseed. Field Crops Research 128: 119–128. doi: 10.1016/j.fcr.2011.12.018.

Ham, G.E., I.E. Liener, S.D. Evans, R.D. Frazier, and W.W. Nelson. 1975. Yield and Composition of Soybean Seed as Affected by N and S Fertilization. Agronomy Journal 67(3): 293–297. doi: 10.2134/agronj1975.00021962006700030004x.

Hitsuda, K., K. Toriyama, G.V. Subbarao, and O. Ito. 2008. Sulfur Management for Soybean Production. Sulfur: A Missing Link between Soils, Crops, and Nutrition. John Wiley & Sons, Ltd. p. 117–142

Hitsuda, K., M. Yamada, and D. Klepker. 2005. Sulfur Requirement of Eight Crops at Early Stages of Growth. Agronomy Journal 97(1): 155–159. doi: 10.2134/agronj2005.0155a.

Jessop, C.E., S. Chakravarthi, R.H. Watkins, and N.J. Bulleid. 2004. Oxidative protein folding in the mammalian endoplasmic reticulum. Biochem. Soc. Trans. 32(Pt 5): 655–658. doi: 10.1042/BST0320655.

Kaiser, D.E., and K.-I. Kim. 2013. Soybean Response to Sulfur Fertilizer Applied as a Broadcast or Starter Using Replicated Strip Trials. Agronomy Journal 105(4): 1189–1198. doi: 10.2134/agronj2013.0023.

Kovalenko, I.V., G.R. Rippke, and C.R. Hurburgh. 2006. Determination of Amino Acid Composition of Soybeans (Glycine max) by Near-Infrared Spectroscopy. J. Agric. Food Chem. 54(10): 3485–3491. doi: 10.1021/jf052570u.

Kuiken, K.A., and M. Lyman. 1949. Essential amino acid composition of soybean meals prepared from twenty strains of soybeans. J Biol Chem (177): 29–36.

Letham, J.L., Q.M. Ketterings, J.H. Cherney, and T.R. Overton. 2021. Impact of sulfur application on soybean yield and quality in New York. Agronomy Journal 113(3): 2858–2871. doi: 10.1002/agj2.20690.

Liu, K. 1997. Soybeans: Chemistry, Technology, and Utilization. Chapman & Hall, New York, NY.

Pazdernik, D.L., A.S. Killam, and J.H. Orf. 1997. Analysis of Amino and Fatty Acid Composition in Soybean Seed, Using Near Infrared Reflectance Spectroscopy. Agronomy Journal 89(4): 679–685. doi: 10.2134/agronj1997.00021962008900040022x.

Riley, N.G., F.J. Zhao, and S.P. McGrath. 2002. Leaching losses of sulphur from different forms of sulphur fertilizers: a field lysimeter study. Soil Use and Management 18(2): 120–126. doi: 10.1111/j.1475-2743.2002.tb00229.x.

Sexton, P.J., S.L. Naeve, N.C. Paek, and R. Shibles. 1998. Sulfur Availability, Cotyledon Nitrogen:Sulfur Ratio, and Relative Abundance of Seed Storage Proteins of Soybean. Crop Science 38(4): cropsci1998.0011183X003800040017x. doi: 10.2135/cropsci1998.0011183X003800040017x.

The American Soybean Association. 2020. U.S. Soybean Meal: Use by Livestock. http://soystats.com/soybean-meal-u-s-use-by-livestock/ (accessed 14 August 2021).

USEPA. 2020. 2020 Power Sector Program - Progress Report.

Wilson, R.F. 2004. Seed Composition. Soybeans: Improvement, Production, and Uses. 3rd ed. ASA, CSSA, SSSA, Madison, WI. p. 621–669

Wu, B., H. Cui, X. Peng, J. Fang, W. Cui, et al. 2012. Effect of Methionine Deficiency on the Thymus and the Subsets and Proliferation of Peripheral Blood T-Cell, and Serum IL-2 Contents in Broilers. Journal of Integrative Agriculture 11(6): 1009–1019. doi: 10.1016/S2095-3119(12)60093-8.

Wylie, L.M., G.W. Robertson, M.G. Macleod, and P.M. Hocking. 2001. Effects of ambient temperature and restricted feeding on the growth of feathers in growing turkeys. Br. Poult. Sci. 42(4): 449–455. doi: 10.1080/00071660120070631.

Zambelli, A. 2021. Current Status of High Oleic Seed Oils in Food Processing. Journal of the American Oil Chemists' Society 98(2): 129–137. doi: 10.1002/aocs.12450.

Tables and Figures

Table 1. ANOVA results and fixed effects significance for amino acid fatty acid data from soybean grown in Chesapeake, Essex, Painter, and Suffolk, Virginia in 2019 and Chesapeake, Eastville, Essex, Painter, Suffolk, and Virginia Beach, Virginia in 2020.

	Whole Model		**Location**	**Year**	**Source**	**Source × Location**
Trait	R-squared	RMSE[a]	df = 5	df = 1	df = 5	df = 25
Cysteine	0.55	0.02	***	***	NS[t]	*
Histidine	0.48	0.02	***	***	NS	*
Isoleucine	0.42	0.04	***	NS	NS	*
Leucine	0.52	0.07	***	NS	NS	*
Lysine	0.44	0.05	***	***	NS	*
Methionine	0.37	0.12	***	NS	*	NS
Phenylalanine	0.48	0.05	***	***	NS	*
Threonine	0.44	0.03	***	*	NS	**
Tryptophan	0.51	0.01	***	***	NS	NS
Valine	0.47	0.04	***	NS	NS	NS
Linolenic acid	0.15	0.10	**	***	NS	NS
Oleic acid	0.49	2.45	***	***	NS	NS
Palmitic acid	0.30	0.65	***	***	NS	NS
Stearic acid	0.33	0.23	***	***	NS	NS

[a]Root Mean Square Error – standard deviation of the residuals

*Significant at the 0.05 probability level **Significant at the 0.01 probability level ***Significant at the 0.001 probability level

[t]NS – not significant

Table 2. Comparison of means between locations for amino acids.

Location	Cysteine	Histidine	Isoleucine	Leucine	Lysine	Methionine	Phenylalanine	Threonine	Tryptophan	Valine
	----------% seed, dry weight basis----------									
Chesapeake	0.56d[a]	1.09a	2.01a	3.22a	2.72a	0.566b	2.18a	1.56a	0.407c	2.08a
Eastville	0.58ab	1.03d	1.91e	2.99d	2.59c	0.556cd	2.00c	1.48c	0.42a	1.95e
Essex	0.56cd	1.06c	1.94d	3.09c	2.62c	0.551d	2.11b	1.53b	0.406c	2.00d
Painter	0.58a	1.08b	1.96c	3.15b	2.67b	0.572a	2.12b	1.56a	0.39d	2.03c
Suffolk	0.52e	1.07b	1.99b	3.17b	2.67b	0.557c	2.13b	1.53b	0.39d	2.05b
Virginia Beach	0.57bc	1.07b	1.98bc	3.14b	2.68b	0.571ab	2.13b	1.54b	0.41b	2.04bc
Grand Mean	0.56	1.07	1.97	3.14	2.67	0.56	2.12	1.53	0.40	2.03
SD[b]	0.03	0.03	0.05	0.10	0.07	0.01	0.07	0.04	0.02	0.06

[a]Within each column, means followed by the same letter are not significantly different.
[b]Standard deviation

Table 3. Comparison of means between locations for fatty acids.

Location	Linolenic acid	Oleic acid	Palmitic acid	Stearic acid
	----------% seed, dry weight basis----------			
Chesapeake	6.55b[a]	23.76b	10.83b	3.66b
Eastville	6.71ab	20.12c	11.36a	3.50c
Essex	6.83a	24.73b	10.80bc	3.53c
Painter	6.87a	26.74a	11.55a	3.74b
Suffolk	6.66ab	23.93b	10.56c	3.53c
Virginia Beach	6.75ab	24.49b	11.41a	3.99a
Grand Mean	6.78	23.97	10.95	3.62
SD[b]	0.59	3.33	0.75	0.27

[a]Within each column, means followed by the same letter are not significantly different.
[b]Standard deviation

Table 4. Comparison of means between years for amino acids.

Year	Cysteine	Histidine	Lysine	Phenylalanine	Threonine	Tryptophan
	----------% seed, dry weight basis----------					
2019	0.58a[a]	1.08a	2.64b	2.10b	1.54a	0.41a
2020	0.55b	1.06b	2.68a	2.12a	1.53b	0.40b

[a]Within each column, means followed by the same letter are not significantly different.

Table 5. Comparison of means between years for fatty acids.

Year	Linolenic acid	Oleic acid	Palmitic acid	Stearic acid
	----------% seed, dry weight basis----------			
2019	6.59b[a]	24.94a	11.32a	3.72a
2020	6.87a	22.99b	10.85b	3.60b

[a]Within each column, means followed by the same letter are not significantly different.

Table 6. Comparison of means sulfur sources for methionine.

Source	Methionine
	% seed, dry weight basis
AMS	0.565a
CS	0.563ab
ATS	0.563ab
UTC	0.562ab
ES	0.560b
Urea	0.560b

[a]Within each column, means followed by the same letter are not significantly different.

Table 7. Results from test slice analysis in JMP for amino acids that had significant interaction between sulfur source and location.

Interaction Slice	Cysteine	Histidine	Isoleucine	Leucine	Lysine	Phenylalanine	Threonine
Chesapeake	*	NS[ɨ]	NS	**	*	*	***
Eastville	NS	NS	NS	NS	NS	NS	NS
Essex	**	NS	NS	NS	NS	NS	NS
Painter	NS	**	*	*	**	*	NS
Suffolk	NS	NS	NS	NS	NS	NS	NS
Virginia Beach	NS	NS	NS	NS	NS	NS	NS
AMS	***	***	***	***	***	***	***
ATS	***	***	***	***	***	***	***
CS	***	***	***	***	***	***	***
ES	***	***	***	***	***	***	***
Urea	***	***	***	***	***	***	***
UTC	***	***	***	***	***	***	***

*Significant at the 0.05 probability level **Significant at the 0.01 probability level ***Significant at the 0.001 probability level
[ɨ]NS – not significant

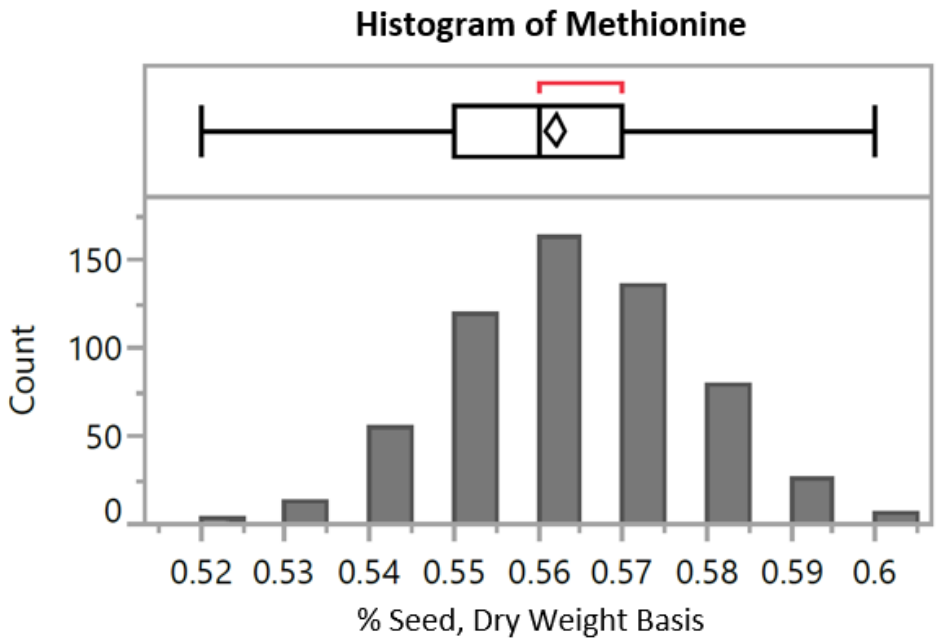

Figure 1. Distribution of methionine observations (n = 588).

Conclusion

The value of harvested soybean is directly impacted by seed composition, especially protein and oil content and their respective monomers. While many resources have been allocated to improve the yield and production efficiency of soybean, great effort is still required to optimize seed composition and quality to meet market demands. In these studies, several approaches were utilized to evaluate methods of genetic improvement for traits of interest and highlighted the importance of management decisions and environmental factors for specific seed components. Results from these studies can be employed in the development of future soybean varieties and as a basis for creating management recommendations.

Using a genome-wide association study, several single nucleotide polymorphisms (SNPs) were identified as being associated with methionine content in soybean seeds. Methionine is the first limiting amino acid and restricts the nutritional value of soybean protein. These SNPs could be used in marker-assisted selection for improving soybean amino acid profiles, and the specific SNPs located in the coding sequence of genes could be the key to understanding physiological mechanisms for methionine content. Genomic selection proved less useful on a genome-wide scale; however, a subset of significant SNPs provided higher predictive accuracy.

A novel phenomic selection method was evaluated using test weight observations. Despite test weight being an important physical seed characteristic and indicator of seed quality, it has been overlooked in soybean breeding programs. Phenomic prediction models displayed highly accurate phenotype estimates and exhibited potential for predicting across environments. This methodology would provide breeders great utility

for test weight, a trait that is difficult to measure in pre-breeding materials and early breeding generations.

The final project sought to understand a factor of environmental effects observed in the previous studies for seed components. Specifically, environmental impact on proteinogenic methionine concentrations, a sulfur containing amino acid, was observed, so influence of sulfur fertilization was selected to investigate. Soybean plots fertilized with granular ammonium sulfate exhibited increased methionine content which is consistent with the plant availability of sulfur sources. By combining this information with new soybean varieties developed for improved seed composition and quality, breeders and producers can partner in the production of economically and nutritionally enhanced soybean.