12-12-2023

Soybean genetics, genomics, and breeding for improving nutritional value and reducing antinutritional traits in food and feed

William M. Singer

Yi-Chen Lee

Zachary Shea

Caio Canella Vieira

Dongho Lee

See next page for additional authors

Authors

William M. Singer, Yi-Chen Lee, Zachary Shea, Caio Canella Vieira, Dongho Lee, Xiaoying Li, Mia Cunicelli, Shaila S. Kadam, Mohammad Amir Waseem Khan, Grover Shannon, M. A. Rouf Mian, Henry T. Nguyen, and Bo Zhang

Received: 21 July 2023 | Accepted: 27 October 2023

DOI: 10.1002/tpg2.20415

The Plant Genome

REVIEW

Special Section: Grain Quality and Nutritional Genomics for Breeding Next Generation Crops

Soybean genetics, genomics, and breeding for improving nutritional value and reducing antinutritional traits in food and feed

William M. Singer[1] | **Yi-Chen Lee**[2] | **Zachary Shea**[1] | **Caio Canella Vieira**[3] | **Dongho Lee**[4] | **Xiaoying Li**[1] | **Mia Cunicelli**[5] | **Shaila S. Kadam**[6] | **Mohammad Aamir Waseem Khan**[6] | **Grover Shannon**[4] | **M. A. Rouf Mian**[5] | **Henry T. Nguyen**[6] | **Bo Zhang**[1]

[1]School of Plant and Environmental Sciences, Virginia Polytechnic Institute and State University, Blacksburg, Virginia, USA

[2]Department of Agriculture, Fort Hays State University, Hays, Kansas, USA

[3]Department of Crop, Soil, and Environmental Sciences, University of Arkansas, Fayetteville, Arkansas, USA

[4]Fisher Delta Research, Extension, and Education Center, University of Missouri, Portageville, Missouri, USA

[5]Soybean and Nitrogen Fixation Research Unit, USDA-ARS, Raleigh, North Carolina, USA

[6]Division of Plant Science and Technology, University of Missouri, Columbia, Missouri, USA

Correspondence
Bo Zhang, School of Plant and Environmental Sciences, Virginia Polytechnic Institute and State University, 185 Ag Quad Lane, Blacksburg, VA 24061, USA.
Email: bozhang@vt.edu

Funding information
United Soybean Board, Grant/Award Numbers: 2332-203-0102, 2333-203-0101

Abstract

Soybean [*Glycine max* (L.) Merr.] is a globally important crop due to its valuable seed composition, versatile feed, food, and industrial end-uses, and consistent genetic gain. Successful genetic gain in soybean has led to widespread adaptation and increased value for producers, processors, and consumers. Specific focus on the nutritional quality of soybean seed composition for food and feed has further elucidated genetic knowledge and bolstered breeding progress. Seed components are historical and current targets for soybean breeders seeking to improve nutritional quality of soybean. This article reviews genetic and genomic foundations for improvement of nutritionally important traits, such as protein and amino acids, oil and fatty acids, carbohydrates, and specific food-grade considerations; discusses the application of advanced breeding technology such as CRISPR/Cas9 in creating seed composition variations; and provides future directions and breeding recommendations regarding soybean seed composition traits.

Abbreviations: ANF, antinutritional factor; BBTI, Bowman–Birk trypsin inhibitor; GMO, genetically modified organism; GWAS, genome-wide association study; HOLL, high oleic, low linolenic; KASPar, kompetitive allele-specific PCR; KTI, Kunitz trypsin inhibitor; LG, linkage group; MAS, marker-assisted selection; ME, metabolizable energy; MG, maturity group; NIRS, near-infrared reflectance spectroscopy; PUFA, polyunsaturated fatty acid; QTL, quantitative trait locus; QTN, quantitative trait nucleotide; RFLP, restriction fragment length polymorphism; RFO, raffinose family oligosaccharide; SAA, sulfur-containing amino acid; SNP, single-nucleotide polymorphism; SSR, simple sequence repeat; TI, trypsin inhibitor; UAV, unmanned aerial vehicle.

1 | INTRODUCTION

Soybean [*Glycine max* (L.) Merr.] is a globally important crop due to its valuable seed components and represents the largest and most concentrated segment of global agricultural trade (Gale et al., 2019). The crop is cultivated on roughly 6% of the world's arable land and has been referred to as the "golden miracle bean" due to its unique seed composition, accounting for approximately 70% of total protein meal and over 60% of total global oilseed production (Hartman et al., 2011; United States Department of Agriculture, 2022; Vieira & Chen, 2021). In 2021, world soybean production totaled 371.7 million metric tons (Mt) with 81.2% of production accounted by Brazil (134.9 Mt), the United States (120.7 Mt), and Argentina (46.2 Mt) (FAO, 2023). The international demand for soybean is driven by versatile feed, food, and industrial end-use purposes provided by a unique seed composition profile. This demand is also highly influenced by China, which purchases 65% of the global soybean supply (De Maria et al., 2020; Gale et al., 2019). Additionally, when compared to other world food crops, soybean had the highest yearly percentage increase in production area over from the 1970s to the 2010s and has continued to grow in global harvested area and production quantities (FAO, 2023; Hartman et al., 2011). Feed and food consumption typically influences the overall production of soybean, while industrial purposes have historically garnered added value through by-products. Soybean seeds are composed of five main seed components: protein, oil, carbohydrates (soluble and insoluble), ash, and water (often displayed as moisture content). Soybean meal (soymeal; protein, carbohydrates, and ash combined) accounts for the majority of seed value through nutritional elements, energy content, and feed conversion, and 1 Mt of soybeans can produce roughly 79,000 kg of meal (USB, 2022; USSEC, 2022). For this reason, most soybeans are crushed to separate the meal from other components, such as oil, to extract the highest value.

Core Ideas

- Soybean genetics are the foundation for improving nutritional composition.
- Soybean genomics can be harnessed to improve nutritional composition.
- Soybean breeding has and will continue to improve nutritional composition.

Soymeal is valuable because of its high utility as a livestock feed ingredient with all nine essential amino acids present: histidine (His), isoleucine (Ile), leucine (Leu), lysine (Lys), methionine (Met), phenylalanine (Phe), threonine (Thr), tryptophan (Trp), and valine (Val) (Phillips, 1993; Qin et al., 2022). This protein quality also provides nutritional value for humans across many cultures. First cultivated in China around 1100 AD, soybean has been a protein staple in many Asian cuisines and remains well-recognized for foods such as tofu and edamame, health benefits, and an increasing interest in plant-based proteins (Chang et al., 2015; Messina, 1999; Messina & Messina, 2010). After its North American introduction in 1765, soybean did not experience large production growth until the early 20th century from a substantial demand increase due to World War II lubricant and oil needs (Chang et al., 2015; Hymowitz & Harlan, 1983). Globalization has provided a platform for international trade that introduced soyfoods into Western cultures and diets, allowing more people worldwide to access soyfoods and to understand their various health and nutritional benefits. In recent years, various soyfood products have surged into the market and can be divided into nonfermented and fermented soyfoods (Figure 1). The versatility of soy-based products has continued to grow alongside technological innovations and is a key to current demand. Subsequently, plant breeders adopted and continue to incorporate emerging methods and technologies such as CRISPR/Cas9 for varietal development to meet fluctuating environmental, economic, and preference-based targets for soybean seed composition.

To assist soybean researchers in expanding toolsets for seed composition improvement, this review seeks to (1) compile traits and genetic foundations for nutritionally added value to seed composition important for feed and food; (2) highlight significant genes, quantitative trait loci (QTLs), and other genetic markers for important seed composition traits; and (3) provide future directions and breeding recommendations regarding seed composition traits in soybean. Additionally, this review may function as a comprehensive knowledge resource for nutritional improvements of soybean seed composition.

2 | PROTEIN AND AMINO ACIDS

2.1 | Introduction to protein and amino acids

Legumes are an excellent source of high-quality protein that contain essential amino acids. Among the legumes, soybean is considered a source of complete protein, containing all nine essential amino acids, and soybean seeds contain approximately 40% protein and 20% oil on a dry weight basis (Banaszkiewicz, 2011). In the western hemisphere, soybean is mainly processed into soymeal for animal feed, as the high concentration of protein (44%–48%) and highly digestible amino acids are an excellent source of feed for the poultry and swine industries. Moreover, soymeal provides amino acids (Lys, Thr, and Tyr) that are deficient in corn, sorghum, and

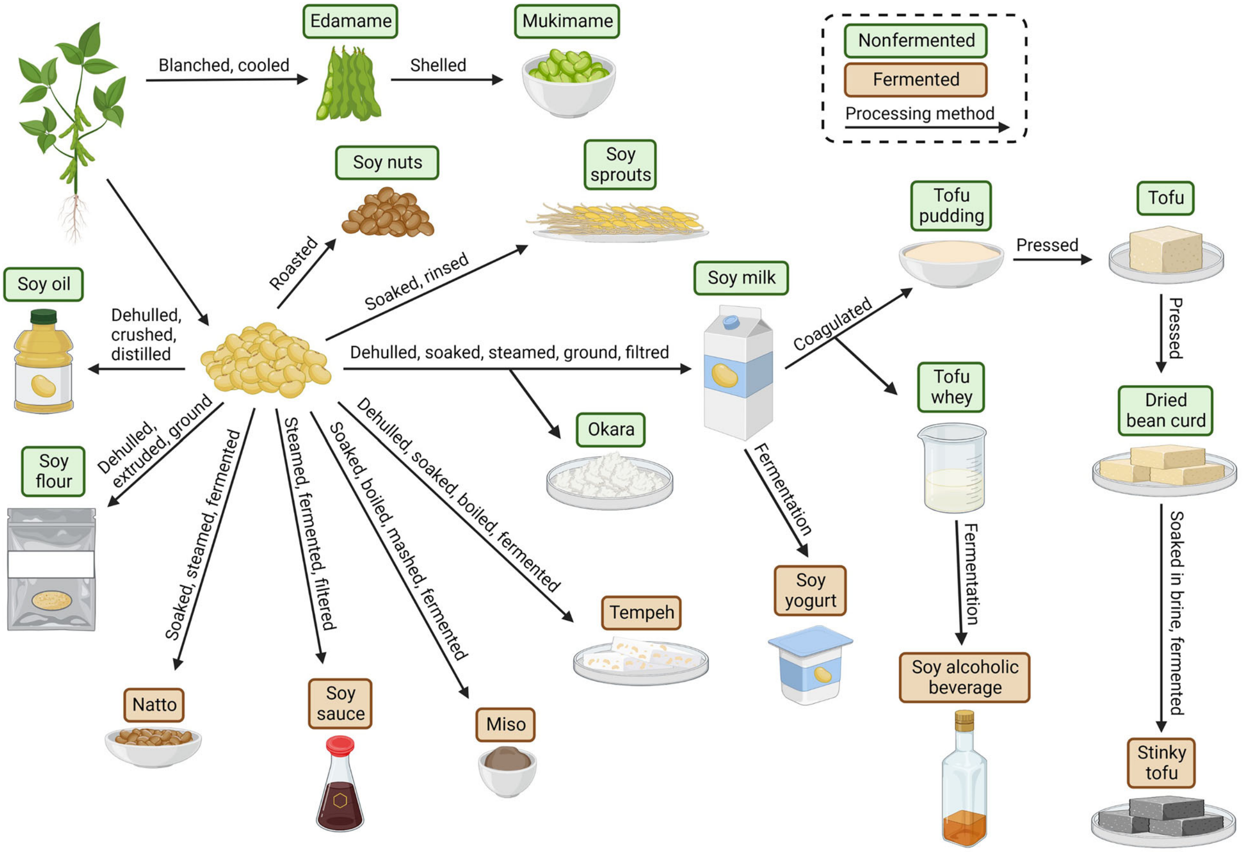

FIGURE 1 Processing methods of fermented and nonfermented soyfoods (created with BioRender.com).

other cereal grains fed to poultry and swine (Cromwell, 1999; Dozier & Hess, 2011; Pettigrew et al., 2002). The US domestic market demands dehulled soymeal with 47%–49% protein concentration at 12% moisture (Bosaz et al., 2019; Guinn, 2002). The protein threshold for producing 48% soymeal protein is approximately 34%–35% seed protein at 13% moisture depending on the oil concentration (USB, 2018).

Protein is required in both human and animal diets to supply essential amino acids. Thus, the blend of amino acids determines the nutritional value of soybean protein (Friedman & Brandon, 2001; Thakur & Hurburgh, 2007; Warrington et al., 2015; Wilcox & Shibles, 2001). Soymeal, which is prepared after extracting oil, is a rich and inexpensive source of proteins with balanced amino acids and is ideal for use in most animal feeds (Krishnan & Jez, 2018). As shown in Figure 2, amino acids have complex and interconnected biological pathways that encourage the direct consumption of protein containing an ideal amino acid profile. Although soybean seed is packed with all nine essential amino acids, three amino acids—Met, Thr, and Lys—are present in reduced quantities (Soy Stats, 2022). Cysteine (Cys) is also deficient, but it is considered "conditionally" essential because animals can convert Met to Cys through the irreversible reaction of cystathionase synthase (Ball et al., 2006). Since monogastric animals such as poultry and swine cannot produce these amino acids naturally, these amino acids must be supplemented in soymeal-based diets with synthetic amino acids at a large expense (~100 million annually) for animal producers (Nill, 2016; Pfarr et al., 2018; Wilcox & Shibles, 2001). Baker et al. (2011) suggest that high-protein soymeal has a greater concentration of amino acids than conventional soymeal. Therefore, high-protein soymeal can provide increased quantities of digestible amino acids, which could solve nutritional deficiency and feeding cost problems.

2.2 | Genetic studies for protein and amino acids

There have been significant efforts made for the identification of protein and amino acid QTLs through biparental conventional mapping as well as genome-wide association studies (GWASs). There are currently 248 QTLs from biparental mapping for protein, including 16 cq-QTLs; cq-QTLs are the confirmed QTLs that are approved by the soybean genetics committee (Grant et al., 2010; Soybase.org, 2022) (Table S1).

The first attempt to map QTLs for high protein was made in 1992 through restriction fragment length polymorphism (RFLP) mapping in a biparental cross developed from a *G. max* experimental line (A81-356022) and *Glycine soja* plant

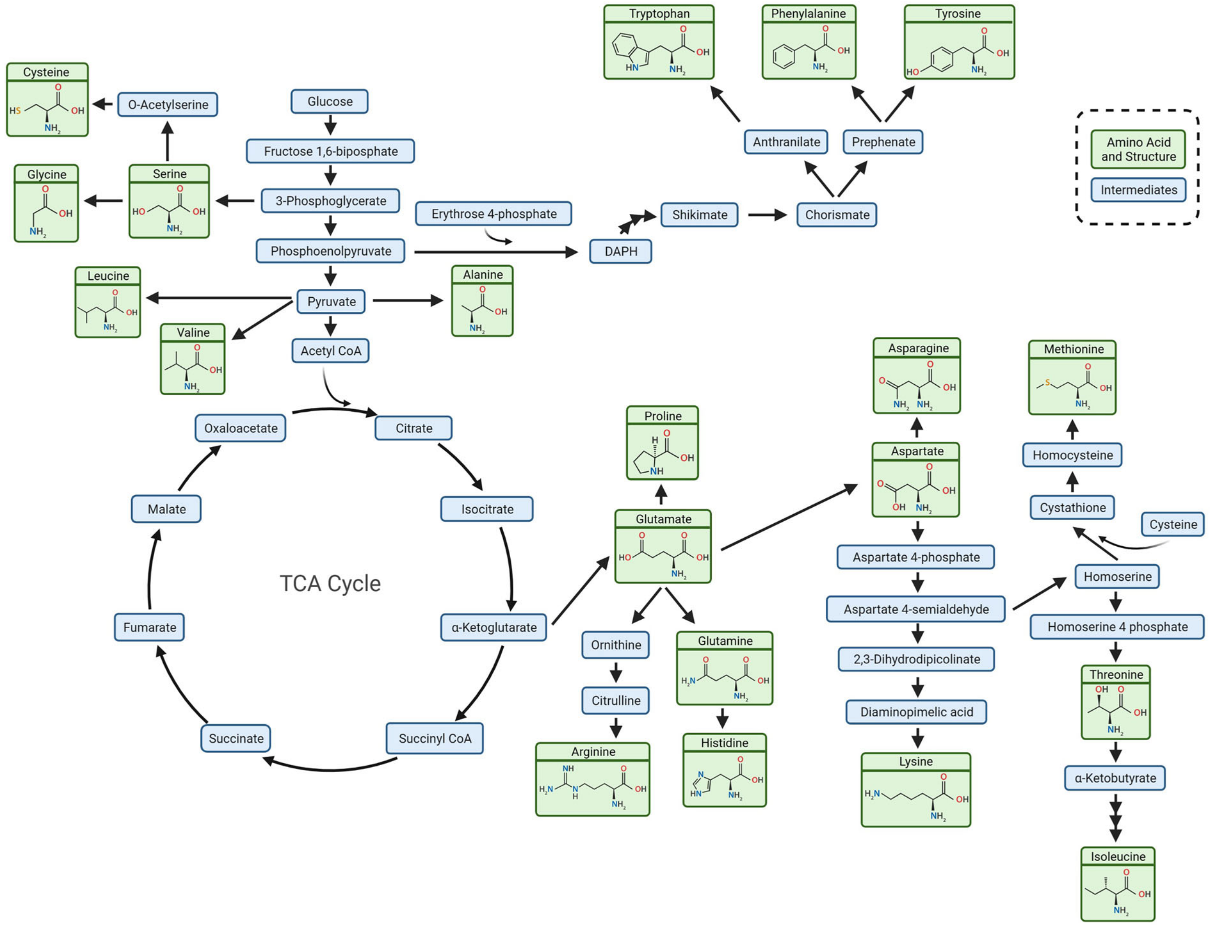

FIGURE 2 The amino acid biosynthesis pathways (created with BioRender.com).

introduction line (PI 468916) (Diers et al., 1992). This study indicated a positive association of high protein with the *G. soja* source. Many subsequent studies for QTL mapping have reported and confirmed chromosome (Chr.) 20 (linkage group [LG] I) and Chr. 15 (LG E) linked to high-protein alleles using different populations and environments (Bolon et al., 2010; Hwang et al., 2014; Jun et al., 2008; Patil et al., 2017; Vaughn et al., 2014; Wang et al., 2015).

Previous studies that used biparental segregating populations have identified QTLs associated with 15 amino acids in soybean seeds (Fallen et al., 2013; Khandaker et al., 2015; Panthee et al., 2006; Warrington et al., 2015). Using 313 diverse soybean germplasm accessions genotyped with a high-density single-nucleotide polymorphism (SNP) array (Song et al., 2013, 2020), Zhang et al. (2018) conducted a GWAS for overall seed composition. In this study, 87 chromosomal regions were identified to be associated with seed composition, explaining 8%–89% of genetic variances. In addition, 54 SNPs, as 92 markers, were associated with 18 amino acids; 38 of the 54 SNPs were associated with only one amino acid, while 11 SNPs were associated with two to 12 amino acids for amino acid concentration in soybean. Hwang et al. (2014) conducted a GWAS on seed protein and oil content that identified 40 SNPs in 17 different genomic regions significantly associated with seed protein. Of these, the five SNPs with the highest associations and seven adjacent SNPs were in the 27.6–30.0 Mbp region of Chr. 20 (LG I). Qin et al. (2019) performed a GWAS of 15 seed amino acid contents on 249 soybean accessions from China, the United States, Japan, and South Korea and reported significant genetic variation.

Yuan et al. (2021) conducted a GWAS for Cys, Met, and total sulfur-containing amino acids (SAAs) in 165 soybean materials genotyped with a high-density SNP array and identified 138 significant SNPs. In this study, a single SNP on Chr. 7 (LG M) was identified in three environments, and *Glyma.07g175700* and *Glyma.07g176000* at the LD of AX-94036794 were identified as candidate genes. Quantitative real-time PCR showed that different expression levels of these genes were observed in high-SAA and low-SAA material, which suggested that these two genes may be involved in SAA synthesis. Singer et al. (2022) conducted a GWAS for proteinogenic Met content in soybean seeds and identified a total of eight significant SNPs associated with higher Met content (five on Chr. 3 [LG N], two on Chr. 8 [LG A2], and one on Chr. 16 [LG J]).

Genotype by environment (G × E) interaction affects the protein quality of soybean, and seed composition even differs across geographical regions within the same country due to environmental variations and genetic factors (Lee et al., 2010; Medic et al., 2014). In the United States, soybeans cultivated in southern states have exhibited higher mean protein and oil concentrations than those in northern states (Breene et al., 1988; Yaklich et al., 2002). Similar results were concluded in other studies that showed soybeans cultivated in the northern and western regions of the Midwest have greater protein loss for each concentration point gain in oil than southern locations (Hurburgh et al., 1990; Piper & Boote, 1999). Additionally, soybean grown in the northwestern states produced soybean seed with lower protein than the southeastern states in the United States (Chung et al., 2003). Multiple factors attribute to regional variation in soybean seed composition, including temperature (Medic et al., 2014; Piper & Boote, 1999; Pipolo et al., 2004), water availability (Rotundo & Westgate, 2009, 2010), and soil fertility (Ham et al., 1975; Nakasathien et al., 2000; Ray et al., 2006).

Interestingly, the effects of temperature on soybean seed content have been inconsistent (Assefa et al., 2019; Patil et al., 2017). Kumar et al. (2006) conducted a multilocation field trials with seven Indian soybean cultivars at four locations, concluding that temperature was positively correlated with seed protein content and negatively correlated with oil content during seed development. In two other studies, warm temperatures (20–28°C) during the soybean growing periods resulted in higher protein content (Sudaric et al., 2006; Vollmann et al., 2000). Several other studies have shown that the increase in temperature resulted in higher oil content while having no effects on seed protein content (Howell & Cartter, 1958; Mourtzinis et al., 2017; Ren et al., 2009). A quadratic relationship between oil and protein concentration with temperature has also been reported (Piper & Boote, 1999; Pipolo et al., 2004).

The composition of soybean seed also differs across global geographical regions. Studies have indicated that Chinese and Brazilian soybean cultivars possess higher protein content than their US counterparts, whereas Argentina exhibits the lowest protein content among the major soybean-producing countries (Grieshop & Fahey, 2001; Karr-Lilienthal et al., 2004; Medic et al., 2014). Ibáñez et al. (2020) conducted a meta-analytic study on soymeal from 18 published papers from 2002 to 2018; they concluded that the soymeal in Brazil exhibited higher levels of crude protein, neutral detergent fiber, and raffinose but lower levels of sucrose, stachyose, and potassium (K) than soymeal from the United States or Argentina. The soymeal from the United States had the most sucrose and stachyose content and the least raffinose and crude fiber among the three countries. The amino acid profile of the crude protein also varied across countries. The Lys, Met, Thr, and Cys concentrations per unit protein were greater in the United States and Argentina soymeal than in the Brazil soymeal.

2.3 | Genomic studies for protein and amino acids

Among all identified QTLs, LG I on Chr. 20 is of major interest due to its close association and large effect with a high protein allele, and three decades after LG I first reporting, *Glyma.20G085100* was reported as a candidate gene for high protein content, contributing to a 2%–3% increase (Fliege et al., 2022). *Glyma.20G085100* has continued to pique interest, and two other groups, Marsh et al. (2022) and Grottel et al. (2022), also reported the same gene by using pangenome information and a transgenic approach, respectively. This gene contains the deletion of 321 bp in the third exon in *G. soja*, which is associated with high protein. However, the insertion of 321 bp in *G. max* alleles alters its function by diverting resources in favor of increasing oil and yield, while decreasing protein. Fliege et al. (2022) reported it as a gain of function allele, while Marsh et al. (2022) and Grottel et al. (2022) mentioned it as a loss of function allele. Though these studies have different opinions about the gain or loss of function allele, it can be argued that the insertion of the 321-bp transposable element disrupts the CCT domain on exon three of this gene, which then reduces the protein. Therefore, it likely should be considered as a loss-of-function allele, in agreement with Marsh et al. (2022) and Grottel et al. (2022). Although this gene is known, the complete functionality of this gene, mechanism of action, and biochemical pathway need further exploration to improve protein content without yield and oil reduction penalties.

The dominant 'Danbaekkong' (Dan) protein allele on Chr. 20 has been linked to higher protein concentration in soybean. This protein allele provides an alternative source of high protein in addition to *G. soja*. Cunicelli et al. (2019) adapted an SNP protocol for protein concentration developed by Warrington et al. (2015) to use marker-assisted selection (MAS) for the Dan allele. They concluded that the lines with the Dan protein allele contained significantly more protein, less oil, and lower yield, which supported the results of previous studies regarding the overall negative correlation between protein and oil. Another study increased protein by overexpressing the Arabidopsis QQS (Qua-Quine Starch; *At3g30720*) orphan gene in soybean, maize, and rice. According to this study, QQS is unique to Arabidopsis, which affects nitrogen and carbon partitioning by interacting with NF-YC transcription factor (Li et al., 2015).

Lee et al. (2019) conducted a comprehensive GWAS to identify QTLs for seed content of protein, oil, and several essential amino acids in soybean. The results indicated that three and five genomic regions were associated with seed protein and oil contents, respectively. In terms of amino acid content, the contents of Cys, Met, Lys, and Thr ($g{\cdot}kg^{-1}$ crude protein) were associated with one, three, one, and four genomic regions, respectively. Interestingly, a QTL on Chr. 5

(LG A1) was reported to increase oil while having no effects on protein content. On the contrary, a QTL on Chr. 10 (LG O) was reported to increase protein content with little effect on oil content, and this QTL co-localized with the maturity gene *E2*/*GmGIa*. The results from this study demonstrated the feasibility of reducing the negative correlation between protein and oil by utilizing trait-specific QTLs.

Malle, Eskandari, et al. (2020) reported QTLs associated with Cys and Met content in a core set of 137 Canadian soybean lines from maturity group (MG) 000–II. Using both a mixed linear model and six multilocus methods with a catalog of 2.18 M SNPs, a total of nine QTLs and 17 quantitative trait nucleotides (QTNs) were identified, of which seven comprise promising candidate genes. Five other candidate genes were identified within four haplotype blocks through multilocus methods. Among these, *Glyma.13g108800*, *Glyma.14g003200*, and *Glyma.14g003400* were annotated to be involved in Cys biosynthesis. Two other candidate genes, *Glyma.04g237300* and *Glyma.16g032200*, had an annotation indirectly associated with Met biosynthesis.

Zhang et al. (2020) mapped a highly effective protein and oil QTLs to a sugar transporter (*GmSWEET39*) gene by a combination of biparental linkage mapping and association analysis using 631 whole-genome sequencing data. *GmSWEET39* was found to have pleiotropic associations with both seed protein and oil, and a small deletion in *GmSWEET39* has been extensively selected and used worldwide for its association with higher seed oil and lower seed protein. Intensive use of this *GmSWEET39* deletion in breeding programs may have resulted in low protein across current soybean cultivars.

2.4 | Breeding efforts and future directions for protein and amino acids

The US soybean seed yield has increased drastically over the past decades due to genetic gains from breeding efforts. However, the strong negative correlation between seed yield and seed protein, and between oil and protein content, has hindered the development of high-yielding cultivars with high protein content (Lee et al., 2010; Patil et al., 2017; Rincker et al., 2014). Rincker et al. (2014) conducted a study to estimate the genetic change in yield and other agronomic traits of northern soybean cultivars MG II–IV in North America over the past 80 years. The study estimated that there was a 23-kg·ha^{-1}·year^{-1} increase in yield, 0.14-g·kg^{-1}·year^{-1} increase in oil concentration, and 0.22-g·kg^{-1}·year^{-1} decrease in seed protein concentration in the MG II cultivars; a 23-kg·ha^{-1}·year^{-1} increase in yield, 0.10-g·kg^{-1}·year^{-1} increase in oil concentration, and 0.22-g·kg^{-1}·year^{-1} decrease in protein concentration in the MG III cultivars; and a 20-kg·ha^{-1}·year^{-1} increase in yield, 0.05-g·kg^{-1}·year^{-1} increase in oil concentration, and 0.16-g·kg^{-1}·year^{-1} decrease in seed protein concentration in the MG IV cultivars. The estimate of annual genetic yield gain was similar to Specht and William (1984), and the estimate of oil and seed protein concentration was similar to Wilcox et al. (1979) and Voldeng et al. (1997). In another study, Rogers et al. (2015) analyzed the genetic change in yield and other agronomic traits of southern soybean cultivars (MG IV–VI) in North America over the past 80 years. The study estimated that the average annual yield increased by 16.8 kg·ha^{-1}·year^{-1}, the oil concentration increased by 0.16 g·kg^{-1}·year^{-1}, and the protein concentration decreased by 0.17 g·kg^{-1}·year^{-1} for the southern soybean cultivars.

While there have been successful attempts at increasing protein concentration in soybean cultivars for decades (Burton et al., 1999; Weber & Fehr, 1970), selection for this trait is hindered by differences in environment (Warrington et al., 2014). Additionally, protein is affected more by genotypic variation than by the environment (Lee et al., 2010; Shorter et al., 1977). In recent years, efforts have been made to increase the competitiveness of US soybean in the global feed market. This led to the development and release of high-protein soybean cultivars and germplasms with high yield potential by public soybean breeding programs in the United States (Bagherzadi et al., 2022; Bhusal et al., 2022; Chen et al., 2022; Fallen et al., 2022; Pantalone & Smallwood, 2018; Pantalone et al., 2017).

Since seed composition traits are controlled by many QTLs, the MAS using only a few major QTLs may not be effective to achieve the ultimate goal of desirable seed compositions . After Meuwissen et al. (2001) proposed the concept of genomic prediction for genomic selection, it was successfully implemented in an animal breeding program for complex quantitative traits (Schaeffer et al., 2006) and subsequently utilized in a plant breeding program (Massman et al., 2013). To date, genomic selection has been successfully applied in soybean breeding programs for key traits, including seed oil and protein (Hemingway et al., 2021; Jarquin et al., 2016; Stewart-Brown et al., 2019), grain yield (Bhat et al., 2022; Ravelombola et al., 2021; Stewart-Brown et al., 2019), agronomic traits (Ma et al., 2016; Zhang et al., 2016), and disease resistance (Bao et al., 2014; de Azevedo Peixoto et al., 2017; Shi et al., 2022). Further nutritional component applications of genomic prediction in soybean will be discussed in later sections.

2.5 | Food grade considerations for protein and amino acids

Soybean breeders are interested in the seed storage proteins glycinin (11S) and β-conglycinin (7S), which account for more than 70% of the total soybean protein content and have been reported to affect the firmness and smoothness of tofu (Guo & Yang, 2015; Meng et al., 2016; Saio & Watanabe, 1978; Zheng et al., 2020). The effects of 11S, 7S, and the 11S/7S protein ratio on tofu texture have been controversial (Cai & Chang, 1999; Meng et al., 2016). Some studies have indicated that 11S and 11S/7S ratio were positively correlated with tofu gel hardness (Cai & Chang, 1999; Chang, 2015; Kang et al., 1991; Kim & Wicker, 2005; Saio, 1979; Saio et al., 1969). Contrarily, some studies have shown that 7S protein resulted in harder gels than the 11S protein and that 11S/7S protein ratio had a negative relationship with tofu hardness (Cai & Chang, 1999; Murphy et al., 1997; Utsumi & Kinsella, 1985). The absence of the 11S globulin polypeptide, 11S A4, was reported to be positively correlated with seed size, tofu hardness, and water holding capacity (James & Yang, 2016; Yang & James, 2014).

The 11S protein is a hexamer of 360 kDa. The six subunits are composed of acidic (A1a, A1b, A2, A3, A4, and A5) and basic (B1a, B1b, B2, B3, and B4) polypeptides linked by disulfide bonds (Bradley et al., 1975; Li & Zhang, 2011; Ma et al., 2010, 2016; Zhang et al., 2021). Multiple genes have been reported to encode for the 11S subunits, and these genes have been designated as Group 1: Gy1 (A1aB2), Gy2 (A2B1a), and Gy3 (A1bB1b); Group 2: Gy4 (A5A4B3) and Gy5 (A3B4); Group 3: two pseudogenes (gy6 and gy8); and Gy7 (polypeptide groups not assigned) (Beilinson et al., 2002; Boehm et al., 2018; Fischer & Goldberg, 1982; Li & Zhang, 2011; Nielsen et al., 1989; Scallon et al., 1985). The 7S is a trimer of α, α', and β subunits that is approximately 180 kDa (Ma et al., 2016; Thanh & Shibasaki, 1978; Zhang, Du et al., 2021). A total of 15 genes (CG1–CG15) have been reported to encode the 7S subunits (Harada et al., 1989; Singh et al., 2015; Yoshino et al., 2002; Zhang et al., 2021).

The 11S globulin was reported to contain three to four times more Met and Cys than 7S, suggesting that 7S ß-subunits are poor in nutritional value Therefore, the nutritional quality of soybean can be increased by the accumulation of 11S globulin while inhibiting the accumulation of 7S globulin in soybean seeds (Ma et al., 2016). To date, more than 100 QTLs associated with soybean seed storage protein have been identified (Grant et al., 2010). Efforts have been made on mapping QTLs that are associated with 7S subunits, and notably, a single dominant gene, Scg-1, was mapped on Chr. 20 (LG I). The gene was reported to suppress the transcription of all 7S subunits without affecting the rest of the plant (Teraishi et al., 2001). In addition, J. Wang et al. (2014) identified a single dominant locus qBSC-1 that controlled the ß-subunits on Chr. 20 (LG I). Regarding the 11S subunit gene family, Beilinson et al. (2002) mapped the Group I Gy1 and Gy3 on Chr. 19 (LG L) and Gy2 on Chr. 3 (LG N). However, the results in Beilinson et al. (2002) were inconsistent with the previous studies where Group I glycinin genes occupied two chromosomal domains (Nielson et al., 1989). Other studies have confirmed that Gy1 and Gy2 are tandemly linked on Chr. 3 (LG N), and Gy3 and Gy7 were mapped on Chr. 19 (LG L) (Cho et al., 1989; Li & Zhang, 2011). In addition, Gy4 was mapped on Chr. 10 (LG O), and Gy5 was mapped on Chr. 13 (LG F) (Chen & Shoemaker, 1998; Diers et al., 1994).

3 | OIL AND FATTY ACIDS

3.1 | Introduction to oil and fatty acids

Soybean plays a multifaceted role in global food security, agricultural production, and the overall economy (Liu et al., 2020; Vieira & Chen, 2021). As the public awareness of eco-friendly and healthy traits in vegetable oils has increased, soybean oil has continuously been highlighted in the biofuel industry and food applications, resulting in a significant increase in global soybean oil production by approximately 44% (42.8–61.6 million Mt) over the past decade (USDA, 2013, 2023). Therefore, soybean breeding programs have devoted tremendous efforts to developing new soybean cultivars producing higher seed oil content (Clemente & Cahoon, 2009; Hartman et al., 2011). Increasing oil content is challenging due to factors such as pleiotropic effects or linkage, leading to a negative correlation with seed protein content (Brummer et al., 1997; Chung et al., 2003; Clemente & Cahoon, 2009; Cober & Voldeng, 2000). In addition, the functionality and quality of soybean oil for food and industrial applications are predominantly determined by its fatty acid profile (Bilyeu et al., 2018; Clemente & Cahoon, 2009). Soybean oil consists of five major fatty acids: palmitic (16:0, ∼13%), stearic (18:0, ∼4%), oleic (18:1, ∼20%), linoleic (18:2, ∼55%), and linolenic acids (18:3, ∼8%) (Bilyeu et al., 2018; Fehr, 2007; Pham et al., 2010). For instance, the oxidative stability and shelf life of soybean oil are substantially reduced by higher seed concentrations of polyunsaturated fatty acids (PUFAs) such as linoleic and linolenic acid (Clemente & Cahoon, 2009; Fehr, 2007). Partial hydrogenation has been widely used to reduce high PUFA concentrations in soybean (Clemente & Cahoon, 2009). However, this process generates trans-fatty acids whose consumption has been associated with negative health consequences, including cardiovascular disease (Clemente & Cahoon, 2009; Danaei et al., 2009; de Souza et al., 2015). It also reduces the lubricity and increases the viscosity of soybean oil, compromising its use in the biodiesel industry (Clemente & Cahoon, 2009; Moser et al., 2007). Soybean oil with high oleic acid (>70%) is also associated with enhanced oxidative stability and nutritional value without the health-threatening trans-fatty acids generated through partial hydrogenation (Bilyeu et al., 2018; Combs & Bilyeu, 2019; Fehr, 2007). Stearic acid content is important not only for oxidative stability but also for the bakery industry because it increases the melting point of fat preventing melting at room temperature (Jeong et al., 2018). Therefore, genetic approaches have proved to be an efficient tool for reducing PUFAs and improving the functionality of soybean oil without negative properties (Clemente & Cahoon, 2009). In this section, the major genetic and genomics advances made in the modification of seed oil content as well as stearic, oleic, and linolenic acid seed contents are reported.

3.2 | Genetic studies for oil and fatty acids

Since the first study reporting QTLs associated with seed oil content (Diers et al., 1992), more than 320 QTLs have been reported across the 20 LGs (Soybase.org, 2022). However, only a few have been detected in multiple genetic backgrounds or environments, and their application in soybean breeding programs through MAS has been limited given the quantitative nature of the trait (Qi et al., 2011). Therefore, only QTLs that have been confirmed by multiple studies are reported in Table S2. Using 60 $F_{2:3}$ lines derived from a cross between the high oil experimental line (A81-356022) and the *G. soja* PI 468916, Diers et al. (1992) identified nine RFLP markers associated with seed oil content across Chrs. 14 (LG B2), 15 (LG E), 19 (LG L), and 20 (LG I). Brummer et al. (1997) were the first to study the stability of oil-related QTLs in soybean. Using over 670 soybean lines derived from eight different populations developed in Minnesota, Indiana, and Nebraska, the study identified 11 RFLP markers distributed in seven chromosomes that were stable across the testing environments (Brummer et al., 1997). Orf et al. (1999) applied simple sequence repeat (SSR) markers in mapping studies related to seed oil content in soybean. By using three biparental populations and over 400 molecular markers, including RFLP and SSR, the study reported two SSR markers (Satt174 and Satt432) associated with seed oil content in Chrs. 5 (LG A1) and 6 (LG C2), respectively (Orf et al., 1999). Many mapping studies based on biparental populations later confirmed similar QTLs using higher density molecular markers and different parental lines (Table S2).

Mutations and targeted downregulation of Δ9 Stearoyl-ACP desaturase (SACPD) have been shown to significantly affect the stearic acid content. Multiple studies have reported the inheritance of the stearic acid trait (Bubeck et al., 1989; Graef et al., 1985, 1988; Pantalone et al., 2002) as well as mapped large genomic regions associated with the regulation of seed stearic acid content (Diers & Shoemaker, 1992; Hyten, Pantalone, Saxton, et al., 2004; Panthee et al., 2006; Reinprecht et al., 2006). Byfield et al. (2006) confirmed the existence of two *SACPD* genes (*SACPD-A* and *SACPD-B*) on Chrs. 7 (LG M) and 2 (LG D1b), respectively (Table S3). P. Zhang et al. (2008) later identified and confirmed a third *SACPD* gene (SACPD-C) on Chr. 14 (LG B2) that encodes a unique isoform of SACPD desaturase responsible for converting stearic acid to oleic acid (Table S3). Recently, other genes of *SACPD-D* and *SACPD-E* were identified on Chr. 13 (LG F), but only *SACPD-D* was reported as a potential new genetic source of high stearic acid using Targeting Induced Local Lesions IN Genomes (TILLING)-by-sequencing technology (Hudson & Hudson, 2021; Lakhssassi et al., 2020) (Table S3). To date, several mutageneses using sodium azide, gamma-ray, and fast neutron in *SACPDs* have achieved stearic acid content ranging from 7% to 28% (Table S4).

The fatty acid desaturase-2 enzyme (*FAD2*) is responsible for the conversion of oleic acid to linoleic acid in developing soybean seeds (Pham et al., 2010; Schlueter et al., 2007). The soybean *FAD2-2* genes consisting of *FAD2-2A* (Schlueter et al., 2007), *FAD2-2B* (Schlueter et al., 2007), and *FAD2-2C* (Bachlava et al., 2009) were found to be widely expressed in the vegetative tissues of the soybean plant (Schlueter et al., 2007), while *FAD2-1A* (Schlueter et al., 2007) is expressed primarily in developing seeds (Tang et al., 2005). Pham et al. (2010) created the high oleic (HO) acid trait in soybean by identifying and combining mutations in *FAD2-1A* and *FAD2-1B*. The study identified three polymorphisms in the *FAD2-1B* alleles of two soybean lines resulting in missense mutation, of which the HO phenotype was observed when combined with existing *FAD2-1A* mutations (Pham et al., 2010). Mutations in the *FAD3-A* (Rennie et al., 1988) gene resulted in a lower linolenic acid (LL) content of approximately 4% (Chappell & Bilyeu, 2007). Combinations of mutations in *FAD3-A* with either *FAD3-B* (Bilyeu et al., 2003) or *FAD3-C* (Bilyeu et al., 2003) lowered linolenic acid to approximately 3%, while combined mutations in *FAD3-A*, *FAD3-B*, and *FAD3-C* resulted in 1% linolenic acid (Bilyeu et al., 2005, 2006, 2011). Later, Pham et al. (2012) combined the *FAD2-1A* and *FAD2-B* mutations with *FAD3* genes resulting in nontransgenic high oleic, low linolenic (HOLL) soybean.

3.3 | Genomic studies for oil and fatty acids

With the availability of high-density SNP markers and advanced statistical methodologies, multiple GWASs identified additional genomic regions associated with seed oil as well as confirmed the previously reported regions from biparental population studies (Bandillo et al., 2015; Cao et al., 2017; Hwang et al., 2014; Lee et al., 2019; Sonah et al., 2015). Hwang et al. (2014) were the first to conduct a GWAS to identify marker–trait associations for seed oil content. The accession panel used in this study included 298 soybean germplasms exhibiting a wide range of seed protein and oil content that were genotyped using 55,159 SNPs (Hwang et al., 2014). The study identified novel genomic regions on Chr. 8 (LG A2) and confirmed previously reported associations on Chr. 20 (LG I) (Chung et al., 2003; Diers et al., 1992). Using a genotyping-by-sequencing (GBS) approach in 304 short-season soybean lines, Sonah et al. (2015) identified eight marker–trait associations that were previously reported across Chrs. 5 (LG A1) (Brummer et al., 1997), 8 (LG A2) (Brummer et al., 1997), 10 (LG O) (Panthee et al., 2005), 14 (LG B2) (Diers et al., 1992), 16 (LG J) (Lee et al., 1996), 19 (LG L) (Diers et al., 1992; Hyten, Pantalone, Sams, et al., 2004; Lee et al., 1996), and 20 (LG I) (Chung et al., 2003; Diers et al., 1992; Hwang et al., 2014). Bandillo et al. (2015) conducted a large GWAS using 12,000 genetically diverse soybean accessions genotyped with the Illumina Infinium SoySNP50K BeadChip. The study narrowed down the previously reported genomic regions on Chrs. 5 (LG A1) (Brummer et al., 1997; Sonah et al., 2015), 15 (LG E) (Diers et al., 1992; Lee et al., 1996), and 20 (LG I) (Chung et al., 2003; Diers et al., 1992; Hwang et al., 2014).

To date, numerous QTLs associated with the five soybean fatty acids have been reported using biparental populations and different accession panels (Table S5). A total of 45 QTLs associated with palmitic acid content were identified across 17 chromosomes. Of those QTLs, seven were identified on Chr. 18 (LG G), and five were identified on Chrs. 5 (LG A1) and 9 (LG K). For stearic acid, 36 QTLs were identified across 17 chromosomes, of which four QTLs were identified on Chrs. 6 (LG C2) and 13 (LG F), and three QTLs were identified on Chrs. 7 (LG M), 14 (LG B2), 16 (LG J), and 18 (LG G). A total of 45 QTLs associated with oleic acid in soybean were reported across 18 chromosomes, of which five QTLs were found on Chr. 18 (LG G) and four QTLs were found on Chr. 5 (LG A1), 13 (LG F), 15 (LG E), and 16 (LG J). Of the 44 linoleic acid-related QTLs across 17 chromosomes, four QTLs were identified on Chr. 3 (LG N), 5 (LG A1), 16 (LG J), and 18 (LG G), and three QTLs each were identified on Chr. 10 (LG O), 13 (LG F), 14 (LG B2), 15 (LG E), and 19 (LG L). A total of 68 linolenic acid-related QTLs have been identified across 18 chromosomes. Of those QTLs, 12 QTLs were identified on Chr. 14 (LG B2), and 11 QTLs were identified on Chr. 15 (LG E).

3.4 | Breeding efforts and future directions for oil and fatty acids

The soybean breeding community has made remarkable progress in developing and releasing cultivars and germplasm with enhanced seed composition, including higher seed oil content and modified fatty acid profiles (Burton et al., 2012; P. Chen et al., 2020, 2021, 2022, 2023; Shannon et al., 2007). Numerous breeding lines originated from first- and second-generation HOLL backcrossing conversions have been used in forward-crossing breeding schemes aiming to develop higher yielding populations with fixed alleles conferring the HOLL phenotype. The strong negative correlation between stearic acid seed content and agronomic traits, such as germination rate and yield performance, remains a bottleneck in breeding for higher stearic acid content. Further research to identify precise and smaller deletions resulting in high stearic content may resolve the negative agronomic traits and facilitate the development of high stearic germplasm and cultivars.

Genome-wide prediction applications were reported to be successful for seed oil content with accuracies ranging from 0.71 (Stewart-Brown et al., 2019) to 0.92 (Jarquin et al., 2016). Although promising, the applicability of genomic prediction for seed oil content is still negligible for many soybean breeding programs given the relatively low cost and high-throughput assessment of seed composition through near-infrared reflectance spectroscopy (NIRS) compared to genotyping costs. Nevertheless, multivariate genomic prediction models leveraging the genetic correlations among traits of interest (Xavier & Habier, 2022) can be explored to improve the identification and selection of genotypes with desirable oil content and/or fatty acid profile and grain yield simultaneously.

3.5 | Food grade considerations for oil and fatty acids

The raw, beany flavor of soybean has been recognized as a major flavor defect when used for human consumption (Mattick & Hand, 1969; Yang et al., 2016). The beany flavor in soybean results from the degradation of PUFAs by lipoxygenase (LOX) into hydroperoxyl derivatives, which further break down into different volatile compounds, resulting in the beany flavor (Wang et al., 2021). Mature soybean seeds contain three LOX isozymes (LOX1, LOX2, and LOX3) encoded by *Glyma.13g347600* (*GmLox1*), *Glyma.13g347500* (*GmLox2*), and *Glyma.15g026300* (*GmLox3*), which are responsible for the formation of the beany flavor. Natural mutants for single, double, and triple lipoxygenase isozymes and the mutated gene fragments have been identified (Kitamura, 1984; Lee et al., 2014; J. Wang et al., 2020). A series of soybean varieties lacking lipoxygenase have been developed using these mutant lines (Chung, 2009; Han et al., 2002). Lenis et al. (2010) reported molecular marker assays designed to distinguish mutant from wild-type alleles for *Lox1*, *Lox2*, and *Lox3*. Wang et al. (2020) adopted a CRISPR/Cas9 strategy targeting the three *GmLox* genes and developed transgenic lipoxygenase-free mutants, which could be used for beany flavor reduction without the use of transgenic techniques.

Besides beany flavor, other off-odors from the oxidation of unsaturated fatty acids during soymilk processing negatively affect consumption. In a GWAS analyzing a natural population of 110 soybean germplasm accessions, nine candidate genes (e.g., *Glyma.06G070100*, *Glyma.07G085200*, and *Glyma.18G017600*) associated with 2-heptenal content in soybean seed (a known cause of soymilk off-odors) were detected on Chrs. 6 (LG C2), 7 (LG M), 10 (LG O), 13 (LG F), 18 (LG G), and 19 (LG L) (Z. Wang et al., 2020). Another GWAS using a natural population of 90 Chinese soybean accessions identified 21 novel QTNs associated with 1-octen-3-ol content in soybean seed, which can cause off-flavors such as mushrooms, lavender, rose, and hay (Xia et al., 2019). These provide important information for selecting soybean cultivars with reduced off-flavor for soymilk.

4 | CARBOHYDRATES

4.1 | Introduction to carbohydrates

Carbohydrates are an important group of storage compounds in soybean seeds. Carbohydrates are found in two forms, soluble and insoluble, of which soluble carbohydrates have garnered more attention than insoluble carbohydrates due to their value in food and feed applications. Typical soybean seeds contain around 15% of soluble carbohydrates, which are divided into three main components: sucrose (~5%), raffinose (~1.5%), and stachyose (~5%) (Hsu et al., 1973; Y. Wang et al., 2014; Wilson, 2004).

Of these three main soluble carbohydrates, sucrose is easily digestible and serves as the main source of metabolizable energy (ME) for animal feeds. Because the ME of sucrose (3900 kcal·kg^{-1}) is significantly greater than that of starch (2918–3396 kcal·kg^{-1}), animal producers are interested in feed formulation with higher sucrose concentration (John, 2008; Ostezan et al., 2023). Additionally, higher sucrose concentration is positively correlated with the sweetness of soy-based products, including natto, tofu, edamame, and soymilk (Rosset et al., 2012; Sui et al., 2020; Wang et al., 2023). Thus, improving sucrose content has become more common in soybean breeding programs due to its higher ME efficiency for animal feed and the natural sweetness in soymeal for human consumption (Ficht et al., 2022; Sui et al., 2020).

On the other hand, the other two soluble carbohydrates, raffinose and stachyose, are considered antinutritional factors (ANFs) known as raffinose family oligosaccharides (RFOs). Raffinose is a trisaccharide composed of a sucrose (disaccharide) and a galactose (monosaccharide), while stachyose is a tetra-saccharide composed of a sucrose and two galactoses (Figure 3). In raffinose synthesis, raffinose synthase (*RS*) initiates chain elongation by adding galactose to sucrose to produce raffinose (Peterbauer & Richter, 2001). Subsequently, stachyose synthase (*SS*) executes the second chain elongation by adding another galactose to raffinose to produce stachyose (Peterbauer & Richter, 2001). RFOs are indigestible in monogastric animals due to the lack of α-galactosidase enzyme activity breaking down the glycosidic linkage between the elongated chains (Chaudhary et al., 2015). Undigested raffinose and stachyose pass to the lower gut as a substrate for microbial fermentation to liberate carbon dioxide, methane, and hydrogen sulfide, which can cause diarrhea and discomfort and eventually reduce feed energy efficiency (Chaudhary et al., 2015; Jo et al., 2018; Kumar et al., 2010). Therefore, removing these ANFs along with improving sucrose in soybean seed is critical to improve ME efficiency for animal feeds as well as for market preference (Coon et al., 1990; Parsons et al., 2000).

4.2 | Genetic studies for carbohydrates

Relative to oil and protein content, a smaller number of QTLs related to soluble carbohydrates (37 QTLs for sucrose and 15 QTLs for oligosaccharides) have been identified through a biparental mapping strategy (Table S6). One of the initial studies to identify QTLs for sucrose content in soybean was reported by Maughan et al. (2000). This study identified 17 QTLs across seven LGs using 149 $F_{2:4}$ individuals from an interspecific cross between a large-seeded breeding line (V71-370) and a *G. soja* line (PI 407162). Kim et al. (2005) reported sucrose- and RFOs-related QTLs on Chrs. 2 (LG D1b), 11 (LG B1), and 19 (LG L), of which two common QTLs on Chrs. 2 (LG D1b) and 19 (LG L) were identified. Later, H. K. Kim et al. (2006) identified QTLs on Chrs. 2 (LG D1b), 6 (LG C2), 12 (LG H), 14 (LG B2), 15 (LG E), 16 (LG J), and 19 (LG L) using the same donor parent (‘Keunolkong’), of which five common QTLs for sucrose and RFOs on Chrs. 12 (LG H), 14 (LG B2), 15 (LG E), and 16 (LG J) were identified. Two $F_{2:3}$ populations were used to identify QTLs associated with reduced stachyose content, where PI 200508 was used as a reduced stachyose donor parent (Skoneczka et al., 2009). This study identified a major QTL on Chr. 6 (LG C2) for sucrose and stachyose content. Another major QTL associated with elevated sucrose and reduced RFOs content was identified on Chr. 11 using a large-seeded breeding line (V71-370) as a donor parent and *G. max* line (PI 87013) (Saghai Maroof & Buss, 2008). Wang et al. (2014) reported three sucrose-related and four RFOs-related QTLs across Chrs. 7 (LG M), 11 (LG B1), 12 (LG H), and 20 (LG I), and a common QTL for sucrose and RFOs content on Chr. 11 (LG B1) was confirmed to be on the same genomic region as identified by Saghai Maroof and Buss (2008). Zeng et al. (2014) identified three QTLs for sucrose content on Chrs. 5 (LG A1), 9 (LG K), and 16 (LG J), accounting for 46%, 10%, and 8%, respectively, of the phenotypic variation. Later, many biparental mapping studies using different donor parental lines and larger marker sets were conducted to locate significant QTLs associated with soluble carbohydrates (Table S6). Recently, researchers have also made tremendous efforts to identify significant SNPs and potential candidate genes for soluble carbohydrates via GWASs using diverse germplasm accessions (Table S7).

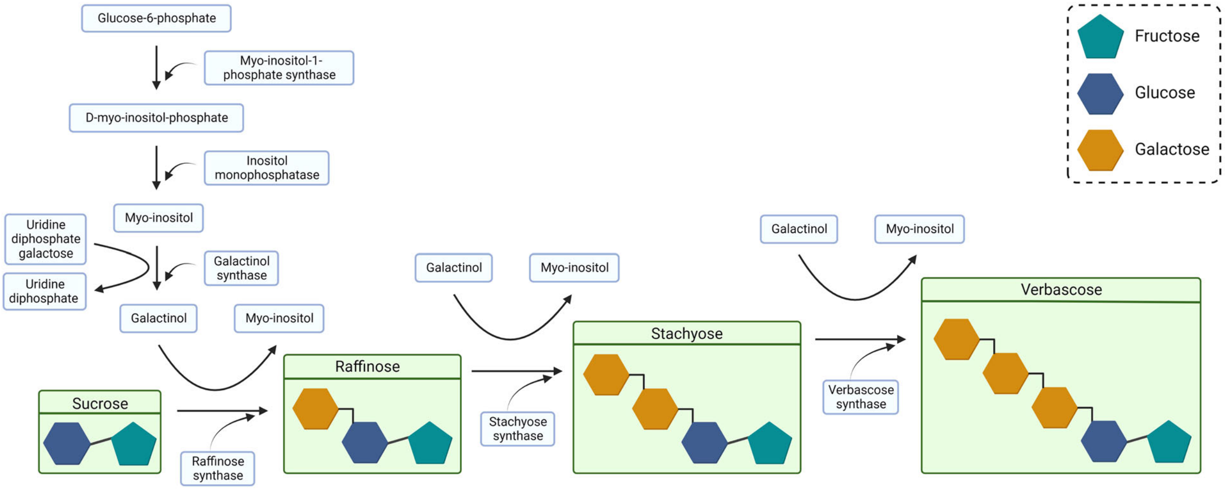

FIGURE 3 The biosynthesis pathways of carbohydrates in soybean (created with BioRender.com).

Environmental effects cannot be ignored despite the significant genetic effects in soybean carbohydrate accumulation. Seed sucrose accumulation is terminated after the maximum seed dry weight, while 70% of RFOs accumulation begins with the onset of seed desiccation and the cease of sucrose accumulation during the pod-filling stages (Fehr et al., 1971; Obendorf, 1997; Obendorf et al., 2009). Due to photoperiod sensitivity, the days of vegetative and reproductive stages in soybean are predominantly determined by day length in the growing area. Therefore, pod-filling stages are closely associated with the designated MG in soybean. Several studies reported a significant correlation between carbohydrate accumulation and soybean growing areas where 11 distinct maturity categories are fitted from north (MG 000) to south (MG VIII) in North America (Scott & Aldrich, 1970; Zhang et al., 2007). Bilyeu and Wiebold (2016) reported that cooler temperature during soybean pod-filling stages had a positive correlation with sucrose accumulation but a negative effect on stachyose content. Furthermore, other studies supported that soybean carbohydrates responded to different growing areas from the northern United States (cooler temperature, early MG) to the southern United States (warmer temperature, late MG) (Bellaloui et al., 2010; Jo et al., 2019; Kumar et al., 2010; Ren et al., 2009). Thus, it is important in soybean carbohydrate research to investigate whether genomic regions discovered by GWAS or QTL mapping overlap soybean maturity genes (Jo et al., 2019).

4.3 | Genomic studies for carbohydrates

As a precursor to many improved carbohydrate contents in soybean, a desirable carbohydrate profile (elevated sucrose and reduced RFOs) was found in PI 200508 in the USDA Soybean Germplasm Collection (Kerr & Sebastian, 2000). Later, a 3-bp deletion within the *RS2* coding region (*Glyma.06g179200*, Wm82.a2.v1) in PI 200508 was characterized and caused one amino acid deletion (Trp) at a highly conserved position 331 (*rs2W331-*) (Dierking & Bilyeu, 2008; Kerr & Sebastian, 2000; Skoneczka et al., 2009). Dierking and Bilyeu (2009) found another significant mutation in the *RS2* coding sequence in an ethyl methanesulfonate (EMS)-mutagenized soybean line using TILLING. This mutation led to one amino acid substitution from Thr to Ile at position 107 (*rs2T107I*) (Dierking & Bilyeu, 2009). Hagely et al. (2013) identified that synonymous *RS3* SNPs (*Glyma05g003900*, Wm82.a2.v1; *rs3snp5/rs3snp6*) played a key role in achieving extremely reduced RFOs in soybean with the presence of *RS2* alleles. Sebastian et al. (2000) developed a soybean mutant line LR33 using chemical mutagenesis, which contained a single base pair polymorphism in a conserved region of D-myo-inositol-3-phosphate synthase 1 (*MIPS1*) gene (*Glyma11g238800*, Wm82.a2.v1). This mutation significantly reduced phytic acid, raffinose, and stachyose and increased sucrose content. Later, Saghai Maroof and Buss (2008) identified a natural mutation in the *MIPS1* coding region, which conferred desirable carbohydrate profiles. To date, three genes, *Glyma05g003900* (*RS3*), *Glyma.06g179200* (*RS2*), and *Glyma11g238800* (*MIPS1*), have been widely used in MAS breeding for desirable carbohydrate profiles.

Valentine et al. (2017) used RNAi silencing to down-regulate the *RS2* gene, which improved the total ME from 2411 to 2703 kcal·kg^{-1}. Cao et al. (2022) used a multiplex CRISPR/Cas9 approach for targeting *RS2* and *RS3* genes. This study generated multiple mutants (double mutant *RS2* + *RS3* and single mutants for *RS2*) and showed significantly reduced raffinose and stachyose content without penalty for plant growth and metabolism. Besides *RS*, galactinol also plays an important role in the biosynthesis of RFOs (Figure 3). A study induced knockouts in two galactinol synthase (*GS*) genes using gRNAs CRISPR/Cas9, of which the total RFOs content was significantly reduced from 64.7 to 42.0 mg·g^{-1} on a dry weight basis (Le et al., 2020). Recently, another study almost eliminated RFOs content in soybean seed via a multiplex genome editing system using three *RS* genes and one *SS* gene simultaneously (Lin et al., 2023). These studies proved the biological functions of *RS*, *SS*, and *GS* genes in the metabolism of RFOs, which are significantly responsible for ME efficiency. However, the bottleneck persists in employing these lines in breeding programs and MAS.

4.4 | Breeding efforts and future directions for carbohydrates

As almost 80% of global soybean production is mainly used for animal feeds, developing new soybean cultivars and germplasms with desirable carbohydrate profiles has been targeted in public and private soybean breeding programs (Hannah & Max, 2021). These efforts have been facilitated to identify promising soybean accessions from the natural germplasm collections (Hou et al., 2009; Kerr & Sebastian, 2000; Kim et al., 2005) or chemically mutated soybean lines (Dierking & Bilyeu, 2009; Hitz et al., 2002). These accessions have been used to identify QTLs associated with the desirable carbohydrate profiles via the QTL mapping strategy. Although many efforts have been made to identify major QTLs and develop molecular markers, only a few cultivars and germplasms have been released (Hagely et al., 2020; Jo et al., 2019; Maughan et al., 2000; Sebastian et al., 2000). The breeding target of new soybean cultivars for the animal feed industry is a sucrose content of over 7% and RFOs content of less than 1%–2% (Pereira, 2020). However, significant environmental effects on carbohydrate profiles have been addressed as a major pitfall in developing new cultivars with stable carbohydrate profiles (Jo et al., 2018, 2019).

Further investigation to identify environmentally stable QTLs will improve breeding efficiency for new soybean cultivars and germplasms with high sucrose and low RFOs content. A negative correlation between protein and sucrose content is another bottleneck in soybean seed (Patil et al., 2017). As high protein content is the most highlighted feature in soybean seed, further investigation is needed to dissect the genetic architecture underlying the negative correlation to maximize animal feed quality.

To the best of our knowledge, there has been only one genomic prediction study for soluble carbohydrate content in soybean, which was conducted by Riaz et al. (2023). With the advent of high-throughput genotyping, genomic prediction would be one of the promising future directions to enhance breeding efficiency in selecting potential new cultivars with desirable soluble carbohydrate profiles.

4.5 | Food grade considerations for carbohydrates

The sweetness flavor of edamame beans is a crucial factor for consumer acceptance, and it is influenced by the content of sucrose and alanine (Ala). Wang et al. (2023) conducted a GWAS of fresh edamame beans of 189 accessions and identified 43 and 25 genetic markers associated with sucrose and Ala content, respectively. Four genes that affect sucrose biosynthesis and 37 novel genes related to sucrose content were identified, along with three genes with likely relevant effects on Ala content and 22 novel Ala-related genes. Moreover, a GWAS of 249 soybean accessions from China, the United States, Japan, and South Korea has found 11 SNP markers related to Ala content in soybean seeds, such as *Gm07_39077446*, *Gm09_43473530*, and *Gm10_12029489* (Qin et al., 2019).

5 | OTHER FUNCTIONAL COMPONENTS

5.1 | Introduction to other functional components

Soybean has many other nutritional components that should be considered during breeding alongside the previously discussed components. This includes ANFs that have a negative effect on humans and animals when consumed and must be removed or inactivated before consumption, such as trypsin inhibitors (TIs), phytate, and allergens. Besides ANFs, soybean also has other beneficial nutritional components that make up a smaller percentage of the nutrient profile, which include isoflavones vitamins, and some minerals (Messina, 1999).

TIs are the most impactful ANFs found in soybean and have two main families, the Kunitz trypsin inhibitor (KTI) and the Bowman–Birk trypsin inhibitor (BBTI) (van den Heuvel, 2021). Both proteins can bind to and inhibit the function of proteases found in the guts of most animals, including fish, cattle, poultry, humans, insects, and others (Liener, 1994). These proteins will inhibit proteases such as trypsin, chymotrypsin, and elastin (Liener, 1994). By inhibiting these enzymes, TIs interfere with protein digestion and prevent animals from obtaining nutrients, specifically proteins, nitrogen, and sulfur, from soymeal. This can cause a wide range of effects, some of which are common across species, while others depend on the species that consumed the soymeal.

Decreased growth is a common effect across species (Flavin, 1982). In certain animals, such as rats, chickens, and mice, TIs can result in pancreatic hypertrophy and reduced feather and fur growth (Hill, 2003). It is worth noting that soybean TIs do not affect humans significantly, especially compared to how they affect other species (Flavin, 1982). This is mainly due to the consumption of cooked soyfood. TIs are inactivated in soymeal by roasting for 30 min at 100˚C or for 22 min at 110˚C, though it is possible to use ultrahigh temperatures, such as 160˚C, to inactivate TIs in a single minute (Avilés-Gaxiola et al., 2018; van der Ven et al., 2005).

Phytate, or phytic acid, is a hexaphosphate of myo-inositol and is considered one of the more significant phosphate reserves for the plant (Jaffe, 1981). Phytate is an ANF due to its capacity to sequester metal ions (such as zinc, calcium, magnesium, and iron), inhibit some enzymes involved with digestion, and decrease the amount of starch and protein absorbed by the animal (Elayabharathy, 2020; Maga, 1982). Unlike TIs, phytate causes similar symptoms across most species.

Soybean ranks as one of the top eight food allergens, and as more soy is consumed globally, soy-based allergic reactions are only expected to become more prevalent (Adolpho, 2019; Wang et al., 2022). The most impactful allergens in soybean include P34, P28, and β-conglycinin (Ogawa et al., 2000; Wilson et al., 2005). Allergic reactions to these allergens include atopic dermatitis, vomiting, diarrhea, difficulty breathing, indigestion, and hives (Adolpho, 2019; Candreva et al., 2016). Exact symptoms and the intensity of the symptoms vary from person to person. There are only a few investigations on allergic reactions of animals to soy products, but some have found dogs, other pets, swine, and calves hypersensitive to soy (Radcliffe et al., 2019). As soybean products become more common, it is expected that there will be more research analyzing any potential allergic reactions to soybean in animal species.

As previously stated in this article, soybean has excellent nutritional value due to widely known seed components. While important, soybean does contain other nutrients of interest, which include isoflavones (genistein, glycitein, and daidzein), vitamins B1, B9, K, and E, and minerals, such as copper, manganese, and molybdenum (Messina, 1999). Table S8 provides information regarding the health benefits of each of the previously mentioned nutrients. It is important to note that these nutrients are present in low/trace levels in the seed tissue (Institute of Medicine, 2001).

5.2 | Genetic and genomic studies for other functional components

Due to TI's significant impact on animal health, developing low-TI soybean varieties is an important goal for soybean breeders. So far, breeders have used naturally occurring mutations in the KTI gene in soybean to develop low-TI varieties. There are five forms of KTI, which are labeled as Tia, Tib, Tic, Tid, and ti, with ti being the low-TI-activity allele (Hymowitz, 1986). The ti allele is recessive and was found to be on Chr. 8 (LG A2) (Cregan et al., 1999). The three SSR markers, Satt409, Satt228, and Satt429, have shown tight linkage with this allele (M. S. Kim et al., 2006). More recently, kompetitive allele-specific PCR (KASPar) markers have been developed in order to help breeders identify low-TI soybean (Rosso et al., 2021).

Due to the function of phytate interfering with animal health, soybean breeders have worked on developing low-phytate lines through a combination of traditional breeding. The line, CX1834-1-2, was determined to have low phytate levels due to two recessive mutant alleles in the soybean homologues for the maize *lpa1* gene, which is responsible for decreased phytate in maize (Gillman et al., 2009). These two alleles were later characterized as *pha1* and *pha2* loci and were found to be associated with Chr. 3 (LG N) and Chr. 19 (LG L), respectively (Gao et al., 2008; Walker et al., 2006). The loci on Chr. 3 (LG N) were found to account for 41% of variation in phytic acid levels, and Chr. 19 (LG L) was found to account for 11% (Walker et al., 2006). Additionally, the SSR markers Satt237 on Chr. 3 (LG N) and Satt527 and Satt561 on Chr. 19 (LG L) were found to be associated with low phytate in soybean seed (Scaboo et al., 2009; Walker et al., 2006). Lastly, mutations in the D-myo-inositol 3-phosphate synthase 1 gene (*MIPS1*) were found to be responsible for improved phosphorus levels in soybean, and breeders were able to develop KASPar markers to identify this mutation (Rosso et al., 2011).

Of the previously mentioned allergens, P34 is the most critical. There are three genes in soybean known to be responsible for P34 expression in soybean seed, which include *Glyma08g12270* (the main contributor), *Glyma08g12280*, and *Glyma05g29130* (Jeong et al., 2013). Breeders have identified two soybean accessions, PI 603570A and PI 567476, that have naturally low P34 levels (Koo et al., 2013). It was later determined that these two accessions have low P34 levels due to an ATGT 4 base pair insertion in the front of the P34 start codon, which is believed to decrease the translation efficiency (Bilyeu et al., 2009; Koo et al., 2013). The low P34 loci was also mapped on Chr. 8 (LG A2) using SSR markers (Watanabe et al., 2017).

While P34 is the main allergen of concern in soybean, β-conglycinin is also important. Researchers have been able to identify two soybean accessions, 'Fukuyutaka' and QY7-25, which have the low β-conglycinin phenotype governed by the dominant gene *Scg-1* (suppressor of β-conglycinin) (Zhang et al., 2021). Furthermore, 10 SNPs were found to be associated with β-conglycinin deficiency between the two lines Fukuyutaka and QY7-25 (Tsubokura et al., 2006). Two

QTLs, linked to Satt461 and Satt249, on Chrs. 17 (LG D2) and 16 (LG J), respectively, were found to be associated with β-conglycinin (Panthee et al., 2004), and a QTL associated with low β-conglycinin was mapped on Chr. 20 (LG I) in the 'Misuzudaizu' and 'Moshidou Gong 503' F_2 mapping population (Tsubokura et al., 2006). Additionally, a *Cgdef* gene was identified and mapped to Chr. 19 (LG L) located between two SSR markers, Satt523 and Sat_388 (Hayashi et al., 2009).

While the other nutrients, like isoflavones and vitamins, are important for human and animal health, they have not received as much attention in terms of breeding research and effort due to them being present in much lower levels than other seed composition traits in soybean. With that being said, isoflavones have had the most progress compared to other nutrients (Wang et al., 2013). Researchers have been able to identify two genomic regions, Satt144–Satt569 and Satt540–Sat_240 (QTIF 1 and QTIM 1, respectively), in soybean that are associated with increased isoflavone content (Han et al., 2016). The QTL mapped on Chr. 13 (LG F) was found to be associated with increased daidzein (QDZF 1), genistein (QGTF 1 and QGTF 2), glycitein (QGCF 1), and total isoflavone content (QTIF 1). The QTL mapped on Chr. 7 (LG M) was found to be associated with increased glycitein (QGCM 1) and total isoflavone content (QTIM 1) (Han et al., 2016; Zeng et al., 2009). Additionally, QTLs mapped on Chrs. 11 (LG B1) and 3 (LG N) were found to be associated with glycitein content, and the QTLs mapped on Chrs. 5 (LG A1) and 3 (LG N) were associated with daidzein content (Meksem et al., 2001). There has been minimal work to determine QTLs associated with vitamin E in soybean, but *Gmpgl1* was found to be responsible for the expression of a thiamine thiazole synthase that allows for thiamin production in soybean (Feng et al., 2019). To date, no QTLs or genes have been discovered that are responsible for folate content in soybean, but it has been identified that soybean germplasm from northern China had higher folate levels compared to germplasms from other areas (Agyenim-Boateng et al., 2022). Multiple QTLs associated with vitamin E in soybean have been identified with three main QTLs located in Chrs. 6 (LG C2) and 2 (LG D1b) (Li et al., 2010). While soybean germplasm evaluation determined that vitamin K1, copper, manganese, and molybdenum can vary across genotypes, no work has been done to identify any genomic regions or genes that are related to the production of these nutrients in soybean (Kastoori Ramamurthy et al., 2014; Thompson et al., 2016).

5.3 | Breeding efforts and future directions for other functional components

Multiple low-TI soybean lines have been developed using the naturally occurring low-TI accession PI 542044 from India, which include NRC101 and NRC102 (Kumar et al., 2013). KTI was removed from multiple elite varieties by utilizing marker-assisted breeding (Kumar et al., 2011; Maranna et al., 2016). Additionally, breeders have harnessed genetic engineering to create low-TI varieties. For example, low-KTI mutant alleles have been created in soybean through CRISPR/Cas9 to selectively knock out KTI expression in seed tissue as well as molecular markers to identify those mutant alleles (Wang et al., 2022). Due to the complex role of TI in plant defense, genetic engineering efforts will likely grow in usage. While low-TI lines have been developed, these lines have low TI levels in all tissues, and as a result, they tend to be more susceptible to pests. An ideotype would be a soybean variety with low-TI content in the seeds while still expressing normal TI levels in other tissues, such as roots and leaves. This would enable the soybean variety to be more resilient against pests while allowing the seed to require less processing time for human and animal consumption. Unfortunately, traditional breeding struggles to develop such a variety due to large gene family affecting TI (Jofuko & Goldberg, 1989). Due to this, genetic engineering tools, including CRISPR/Cas9, could be useful for breeding low-TI soybean varieties due to their ability to selectively target and knock out specific genes. While both KTI and BBTI are important in soybean, KTIs have received more attention than BBTI due to higher concentration in soybean seeds. BBTI has started to gain more attention in breeding because its high activity level has significant effects on animals when consumed (Z. Chen et al., 2020). Because of this, it is expected that breeding efforts will begin for developing varieties that have low BBTI content and activity.

Breeders have used low-phytic-acid mutants, *MIPS1* and *IPK1*, to cross breed with cultivars containing normal phytate levels (Goßner et al., 2019). This study has been able to lead to the development of soybean varieties with phytate levels reduced by 40%–70% depending on which parents were used (Goßner et al., 2019). Besides this, other breeding attempts for reduced phytic acid levels have resulted in soybean varieties with 50%–95% reduced phytate level (Raboy, 2002). Although phytate is an important ANF, it has not received as much attention in breeding as other ANFs, such as TI. This is most likely due to the inactivation methods of phytate being cheap and easy to use (Lei & Porres, 2011). Currently, phytate can be inactivated by adding the enzyme, phytase (Lei & Porres, 2011). In addition to phytase being cheap and easy to use, it is effective at improving the availability of trace elements and minerals, such as calcium, zinc, and iron (Hurrell, 2003; Lei et al., 1993). It is expected that low-phytate breeding in soybean will continue to some degree but not to the same extent as low-TI breeding largely due to phytase efficacy.

Due to P34's relevance in human and animal health, soybean breeders have started screening for low-P34 lines. So far, seven *G. soja* accessions and two *G. max* accessions (PI 603570A and PI 567476) appear to have low levels of the

P34 protein (Guan et al., 2012). These two accessions have already been used to develop low-P34 soybean lines, and breeding efforts for low P34 will likely increase. Regarding P28, null accessions were identified in north spring soybean, Huanghuai summer soybean, and south China soybean (Guan et al., 2012). Additionally, roughly 80% of Japanese cultivars were found not to contain the P28 protein (Bando et al., 1996). Due to these accessions and cultivars lacking P28, they will most likely be incorporated into breeding schemes to help develop low P28 lines. While soybean mutant lines that lack or have reduced β-conglycinin levels have been developed, they showed poor plant growth (Guan et al., 2012). While the reason is not understood, these mutants can be used in breeding schemes and crossed with elite varieties to combine low β-conglycinin, improved plant growth, and enhanced yield compared to the mutant lines. Besides traditional breeding, a transgenic approach can also be utilized. Transgenic soybean lines that lack P34 have been created through transgene-induced gene silencing that had comparable growth and yield to the wild type (Herman et al., 2003). Due to the success of genetic modification to develop a soybean line with no P34, it is possible that a transgenic approach will also be used to develop soybean lines that lack other allergens. Ideally, breeders would develop a soybean line that lacks all soybean allergens, and genetic modification may be one of the best methods to accomplish this. While breeding for low-allergenic soybean lines is still relatively new, breeders do have the necessary tools to develop high-yielding, low-allergenic lines.

While there are other important nutritional factors in soybean, such as isoflavones, vitamins, and minerals, they have received minimal attention in breeding. This is not surprising given that other traits in soybean tend to be more significant and have a higher economic impact, like protein, oil, and ANF content. However, there has been some work to breed for an increase in isoflavone content in soybean. Some researchers crossed the low-isoflavone-content NS-L-146 variety to three other lines, NS Zenit, NS Maximus, and NS Virtus, which exhibited low, medium, and high isoflavone content, respectively (Miladinović et al., 2019). This work surprisingly found that the cross between NS-L-146 and NS Maximus, the medium isoflavone line, resulted in offspring with the highest isoflavone content (Miladinović et al., 2019). Since high-isoflavone soybean lines were identified, it is possible for breeders to use these lines to increase isoflavone content. Some focus will likely shift to other vitamins and minerals. This will include efforts to gain a better understanding of underlying genetics as well as identify naturally occurring soybean accessions that have high levels of these vitamins and minerals. Additionally, as genetic understanding is unlocked, new technologies, such as CRISPR/Cas9, can be used to develop soy varieties that have increased levels of these nutrients. As soybean consumption continues to increase, it is expected that breeders will start to increase their focus on other nutrients in addition to major soybean seed characteristics in order to continue improving soybean for consumer health.

5.4 | Food grade considerations for other functional components

Raw and roasted soy flour are rich in calcium (from 188 to 206 $\text{mg}\cdot 100\cdot\text{g}^{-1}$) (Martino et al., 2008). However, the positive correlation between calcium content in the seed coat and seed hardness poses a challenge during processing, especially in soybean sprouting, natto manufacturing, and fermented soy products, due to the occurrence of stone seeds that do not absorb water during soaking (Mullin & Xu, 2001; Saio, 1976). However, this correlation is not always consistent due to environmental effects such as soil type and temperature variations (Chen et al., 2001; Mullin & Xu, 2001).

Using soybean varieties with low calcium content can partially solve the stone seed issue. Soybean seed calcium content varies between cultivars (0.19%–0.52%) and has a heritability of 41%–63% (Chen et al., 2001; Malle, Morrison, et al., 2020; B. Zhang et al., 2008). Zhang et al. (2009) reported four QTLs associated with calcium content in soybean seeds on Chrs. 7 (LG M), 8 (LG A2), and 20 (LG I), using 148 SSR markers and 178 $F_{2:3}$ and 157 $F_{2:4}$ lines. Orazaly et al. (2018) reported markers Satt267 and Sat_345 on Chr. 1 (LG D1a), Sat_288 on Chr. 7 (LG M), Sat_228, Satt341, and Sat_392 on Chr. 8 (LG A2), Satt547 on Chr. 16 (LG J), and Satt002 on Chr. 17 (LG D2) as reliable for calcium content selection. Additionally, Malle, Morrison, et al. (2020) conducted a GWAS based on a core set of 137 Canadian soybean lines. Two QTLs, Ca_#3 (Gm06: 3,354,869) and Ca_#4 (Gm09: 6,092,970), were associated with soybean seed calcium content, and a candidate gene Glyma.06G046000, involved in calcium transport, was found to be significantly associated with calcium content in seeds and expressed in various plant tissues (i.e., young leaves, flowers, main roots, pods as well as in seeds).

6 | EMERGING TECHNOLOGIES

6.1 | Introduction to emerging breeding technologies

Throughout this article, the various seed composition traits of soybean as well as breeding efforts to improve them have been discussed. While traditional breeding has accomplished a significant amount of progress, there are new, emerging technologies that breeders have started to incorporate. These technologies include CRISPR/Cas9 gene editing, high-throughput phenotyping, and high-dimensional genomic data for genomic prediction and selection purposes. These methods have been employed to make varieties with unique traits, overcome obstacles that traditional breeding has struggled against, and increase the speed and accuracy of breeding.

CRISPR/Cas9 allows scientists and breeders to precisely edit the DNA of living cells. It can be utilized in many organisms including plants, animals, bacteria, and fungi. The technology works by using a "guide" RNA molecule to direct a Cas9 enzyme to a specific location in the genome, where it makes a cut to trigger the cell's natural DNA repair mechanisms, which can be manipulated to either introduce specific genetic changes or remove existing ones (McCarty, 2020). This creates a powerful instrument for quickly generating genetic diversity for breeding, allowing researchers to develop new varieties with unique traits (Synthego, n.d.). Many successful applications have already reported increased yield, nutritional content, and crop resilience to biotic and abiotic stressors (Zaidi et al., 2020). For example, CRISPR/Cas9 has improved yield in rice by targeting the Gn1a, DEP1, GS3, and IPA1 genes (Li et al., 2016), increased vitamin A content in rice (Paine et al., 2005), developed powdery mildew-resistant tomatoes (Nekrasov et al., 2017), and developed drought-resistant maize varieties (Shi et al., 2017). Overall, CRISPR/Cas9 is an incredibly versatile tool that has the potential to revolutionize crop breeding and help to address global food security challenges.

High-throughput phenotyping involves the use of an automated system to analyze quickly and precisely one or several traits. Due to their ability to collect large-scale and high-dimensional data, high-throughput phenotyping platforms have gained the attention of plant breeders as they can reduce the time, cost, and labor of data collection in breeding programs. Examples include NIRS for measuring seed composition traits as well as unmanned aerial vehicles and ground robots for measuring a variety of traits, ranging from agronomic traits to stress resilience (Cabrera-Bosquet et al., 2012; Yang et al., 2020).

The availability of high-dimensional genomic data (Song et al., 2013, 2020) combined with advancements in genome-based prediction and selection has contributed to accelerated genetic gains in plant breeding programs (Jarquín, Crossa, et al., 2014; Jarquín, Kocak, et al., 2014; Vieira et al., 2022; Widener et al., 2021). The concept revolves around using all known molecular marker information to develop prediction models for the phenotype of interest with unobserved genotypes (Meuwissen et al., 2001). Thus, it allows for earlier identification and selection of promising genotypes in the breeding pipeline, which not only reduces costs, time, and space but potentially enhances genetic gain by shortening breeding cycle length and increasing selection intensity (Crossa et al., 2017; Jarquín, Kocak, et al., 2014; Vieira & Chen, 2021; Wartha & Lorenz, 2021).

6.2 | Emerging technology uses in soybean breeding for seed composition

CRISPR/Cas9 has been used to edit a variety of traits in soybean including biotic and abiotic stressors tolerance, agronomic traits, and improved seed composition and nutritional profile (Xu et al., 2020). For instance, CRISPR/Cas9 has been used to create early-flowering soybean genotypes, drought-tolerant soybean, and *Phytophthora sojae*-resistant soybean (Han et al., 2019; Liu et al., 2023; Zhong et al., 2022). While there has been a significant amount of work involving the use of CRISPR/Cas9 to augment a wide range of traits, this section will only focus on CRISPR/Cas9's involvement with traits related to seed composition. This technology has been used to increase the content of nutrients as well as decrease the content of ANFs. It has been used to increase the oleic acid content while simultaneously decreasing the linolenic acid (Chen et al., 2011; Demorest et al., 2016; Wu et al., 2020), as well as combining these traits with a decrease in linoleic acid (Do et al., 2019). CRISPR/Cas9 has also been utilized to improve protein content through the editing of the *AIP2* gene (Shen et al., 2022). Remarkably, there have been cases of developing soybean varieties combining both improved protein and oil content, which were previously demonstrated to have a strong pleiotropy and negative correlation. In this particular case, researchers were able to develop soybean progenies with increased oleic acid and protein contents while significantly decreasing linolenic acid content by knocking out the *GmFAD2-1A* and *GmFAD2-2A* genes (Wu et al., 2020). This multiplexing of traits has also been used to improve the isoflavone content of soybean and decrease the content of the RFOs (Cao et al., 2022; Zhang et al., 2020). Lastly, CRISPR/Cas9 has been used to significantly decrease the content of ANFs, KTI, P28, and P34 of soybean seeds (Sugano et al., 2020; Wang et al., 2022).

While high-throughput phenotyping is already an extremely useful tool for soybean breeders, improved methods are constantly being developed and adopted by the soybean breeding community. NIRS has substantially contributed to the improvements of protein and oil in soybean. The development and calibration of superior models have allowed NIRS to estimate additional seed composition traits, such as amino acids, fatty acids, sucrose, and even total isoflavone content (Choung, 2010; Pazdernik et al., 1997; Sato et al., 2008). The advancements in robotics, artificial intelligence, and NIRS capacities are allowing the development of real-time, nondestructive, single-seed characterization for physical and chemical properties. An example is the QSorter Explorer, which accurately sorts and groups seeds on an individual basis based on a variety of traits including protein and oil as well as specific fatty acids (QSorter Technology, n.d.). This allows breeders to easily identify seeds with desirable composition as well as facilitating the process of purifying seeds for field trials and commercialization.

Genomic selection has been used in breeding schemes to enhance different traits in soybean, including stress resilience, seed composition, agronomic traits, and ultimately seed yield. For protein and oil content, genomic selection has been found to have moderate to high efficacy and accuracy, ranging from 40% to 80% (Duhnen et al., 2017; Jarquin et al., 2016; Smallwood et al., 2019; Stewart-Brown et al., 2019). While protein and oil content are two of the primary traits studied in soybean breeding, genomic prediction and selection pose limited benefits given the availability of methods that can effectively quantify seed protein and oil quickly and cost effectively. Instead, genomic prediction and selection have focused on traits that are labor intensive and/or costly to measure, such as amino acids and carbohydrate profiles. Genomic prediction and selection for amino acid composition are effective and have high selection efficiency with up to 16 amino acids including Ala, His, Ile, Leu, Lys, Phe, Thr, Met, and Val as well as arginine, asparagine, glutamine, glycine, proline, serine, and tyrosine—ranging between 20% and 80% depending on the amino acid (Qin et al., 2019; Singer et al., 2022). Additionally, multivariate genomic prediction models can be deployed to improve multiple traits simultaneously (Jia & Jannink, 2012). Multivariate models have been used to predict traits of interest, such as chlorophyll content in conjunction with soybean cyst nematode (*Heterodera glycines*) resistance, and have been found to improve prediction accuracies compared to univariate models (Okeke et al., 2017; Ravelombola et al., 2019). In seed composition, multivariate models have the potential of improving a variety of seed composition traits simultaneously, including selection for higher protein and oil content, as well as amino acid, carbohydrate, and ANF content.

6.3 | Future directions for emerging technologies

Although there have been many successes in soybean seed composition improvement, further advancements have been hindered by technological obstacles. The two primary obstacles are obtaining stable soybean mutants after transformation and finding suitable genotypes of soybean that can be transformed (Xu et al., 2022). As of now, nonelite cultivars, such as Jack and Williams 82, are primarily used in transformation (Paz et al., 2004). While these lines are highly suitable for transformation, they lack the high-yielding genetic background of modern cultivars. Currently, few elite lines have been transformed (Jing et al., 2016; Zhang et al., 2014). Ideally, breeders would be able to transform a wide variety of soybean varieties to develop elite lines with improved nutrient traits. To fully utilize CRISPR/Cas9 to accelerate breeding, more work is needed to overcome these obstacles. Doing this will enable breeders to transform elite lines that are adapted to their respective areas of interest. This will increase the breeding speed by allowing breeders to eliminate the step of crossing the transformed cultivars with adapted, elite lines. Additionally, CRISPR/Cas9 has been successfully used to multiplex traits. As gene editing technology continues to improve, there is no doubt that breeders will use it to multiplex a variety of improved seed composition traits into one variety. This will likely expand to multiple composition traits, such as a variety having allergens, TI, and other ANFs knocked out.

Similarly, the high-throughput phenotyping methods will continue to advance. Even though NIRS has been heavily used to quickly measure a variety of seed composition traits, there are still some traits that it cannot precisely measure. These traits include ANFs, such as KTI and BBTI, vitamin and mineral content, carbohydrates, and individual isoflavones. As this technology continues to improve, breeders will be able to select desirable traits in a high-throughput manner with minimal labor requirements, costs, and inputs.

Genomic prediction and selection have revolutionized plant breeding by predicting unobserved traits that are labor intensive and costly to measure. Although minimal applications have been deployed in seed composition-related traits, genomic prediction and selection can substantially help improve seed composition traits such as ANFs, isoflavones, vitamins, minerals, and carbohydrates. For instance, genomic prediction and selection can be used for KTI and BBTI as opposed to just overall TI content. Lastly, due to the increased efficacy of multivariate models compared to univariate models, genomic selection can be used to help breeders improve multiple traits simultaneously. This could be particularly useful for breeding for allergens and isoflavones since it is necessary to have soybean varieties with augmented concentration of many allergens and isoflavones simultaneously.

7 | CONCLUSION

Soybean is a historically and contemporarily important crop because of its valuable seed composition and the ability of plant breeders to adapt soybean globally in addition to maximizing seed value. Diverse and innovative processing methods have also expanded the functionality of single seed by unlocking each unique seed component. This utility provides measurable, added value that can be obtained by targeting specific market demands through traditional, molecular, and novel breeding approaches. More specifically, breeders can continue to use genetic and phenotyping technologies to emphasize seed composition alongside agronomic traits.

With the advent of MAS, plant breeders could now successfully use genetic information to accelerate their breeding programs through known genes or genetic linkage that was strong enough between a marker and an unknown causal variant. As markers improved alongside next-generation sequencing, high-density molecular markers, and whole-genome resequencing, the return on investment for MAS also improved. These advancements have also bolstered the progress of genomic prediction and genomic selection models. Application of genomic prediction and genomic selection to many quantitative seed composition traits could discover genetic gain previously limited by MAS.

Similar progress has occurred in technologies for genetic engineering and high-throughput phenotyping, which could further enhance a plant breeder's ability to develop new varieties. From genetically modified organisms (GMOs) to gene editing, breeding schemes have the potential to shift paradigms from selecting the best to designing the best. In soybean, genetic engineering shows enormous potential for translating genetic mechanisms to field realization, with GMO varieties finding significant herbicide tolerance success and gene editing methodologies evolving rapidly. However, neither has shown an immediate proclivity for nutritional improvement. A combination of seed compositions being broad, quantitative traits and the soybean plant itself being difficult to transform will require substantial efforts to overcome. At the other end of the spectrum, high-throughput phenotyping technologies could allow breeders to surmount the phenotyping bottleneck by quantifying phenotypes with greater breadth, depth, and speed. Specifically, sensors and calibration models that could quantify seed components would make nutritional breeding targets more accurate and dependable, and partitioning out variations over time or in combination with environmental conditions will be able to investigate complex inheritance and genetics by environment mechanisms. This will be increasingly important as breeders tackle imminent challenges such as selecting for dynamic environments resulting from shifting climates and erratic weather patterns, combining comprehensive nutritional compositions or <50% protein in high yielding backgrounds, and fulfilling the growing demand for plant-based protein sources.

Soybean breeders have aptly adopted technologies to match producer, environmental, and market needs of soybean to develop consistent and competitive varieties. Continually maximizing value and utility through seed composition will require the same tactical application of emerging technologies to the diversity of soybean germplasm and the complexities of environmental interactions.

AUTHOR CONTRIBUTIONS

William M. Singer: Conceptualization; writing—original draft; writing—review and editing. **Yi-Chen Lee**: Conceptualization; writing—original draft; writing—review and editing. **Zachary Shea**: Conceptualization; writing—original draft; writing—review and editing. **Caio Canella Vieira**: Conceptualization; writing—original draft; writing—review and editing. **Dongho Lee**: Conceptualization; writing—original draft; writing—review and editing. **Xiaoying Li**: Conceptualization; writing—original draft; writing—review and editing. **Mia Cunicelli**: Conceptualization; writing—original draft; writing—review and editing. **Shaila S. Kadam**: Conceptualization; writing—original draft; writing—review and editing. **Mohammad Aamir Waseem Khan**: Conceptualization; writing—original draft; writing—review and editing. **Grover Shannon**: Conceptualization; writing—review and editing. **M. A. Rouf Mian**: Conceptualization; writing—review and editing. **Henry T. Nguyen**: Conceptualization; writing—review and editing. **Bo Zhang**: Conceptualization; writing—review and editing.

ACKNOWLEDGMENTS

The authors would like to dedicate this review to Dr. Pengyin Chen for his contribution to soybean breeding and seed composition research. Support for activities contributing to this review was provided by United Soybean Board Grant Nos. 2333-203-0101 and 2332-203-0102. The former was funded in partnership with the Foundation for Food and Agriculture Research.

CONFLICT OF INTEREST STATEMENT

The authors declare no conflicts of interest.

DATA AVAILABILITY STATEMENT

There are no original data associated with this article. Referenced data are available in the literature.

ORCID

William M. Singer https://orcid.org/0000-0002-6149-6797
Mohammad Aamir Waseem Khan https://orcid.org/0000-0001-9734-8123
Henry T. Nguyen https://orcid.org/0000-0002-7597-1800
Bo Zhang https://orcid.org/0000-0002-0155-385X

REFERENCES

Adolpho, N. (2019). *Soy allergy: Causes, symptoms & treatment*. https://acaai.org/allergies/allergic-conditions/food/soy/

Agyenim-Boateng, K. G., Zhang, S., Islam, M. S., Gu, Y., Li, B., Azam, M., Abdelghany, A. M., Qi, J., Ghosh, S., Shaibu, A. S., Gebregziabher, B. S., Feng, Y., Li, J., Li, Y., Zhang, C., Qiu, L., Liu, Z., Liang, Q., & Sun, J. (2022). Profiling of naturally occurring folates in a diverse soybean germplasm by HPLC-MS/MS. *Food Chemistry*, *384*. https://doi.org/10.1016/j.foodchem.2022.132520

Assefa, Y., Purcell, L. C., Salmeron, M., Naeve, S., Casteel, S. N., Kovács, P., Archontoulis, S., Licht, M., Below, F., Kandel, H., Lindsey, L. E., Gaska, J., Conley, S., Shapiro, C., Orlowski, J. M., Golden, B. R., Kaur, G., Singh, M., Thelen, K., … Ciampitti, I. A. (2019). Assessing variation in US soybean seed composition (protein and oil). *Frontiers in Plant Science*, *10*, 298. https://doi.org/10.3389/fpls.2019.00298

Avilés-Gaxiola, S., Chuck-Hernández, C., & Serna Saldívar, S. O. (2018). Inactivation methods of trypsin inhibitor in legumes: A review. *Journal of Food Science*, *83*, 17–29. https://doi.org/10.1111/1750-3841.13985

Bachlava, E., Dewey, R. E., Burton, J. W., & Cardinal, A. J. (2009). Mapping candidate genes for oleate biosynthesis and their association with unsaturated fatty acid seed content in soybean. *Molecular Breeding*, *23*(2), 337–347. https://doi.org/10.1007/s11032-008-9246-7

Bagherzadi, L., Gillen, A. M., McNeece, B. T., Mian, R., Song, Q., Taliercico, E., Fallen, B., Li, Z., & Carter, T. E. (2022). Registration of USDA-N7005 soybean germplasm with high yield and 62.5% pedigree from Japanese accessions Tamahikari and PI 416937. *Journal of Plant Registrations*, *16*, 641–648. https://doi.org/10.1002/plr2.20209

Baker, K. M., Utterback, P. L., Parsons, C. M., & Stein, H. H. (2011). Nutritional value of soybean meal produced from conventional, high-protein, or low-oligosaccharide varieties of soybeans and fed to broiler chicks. *Poultry Science*, *90*, 390–395. https://doi.org/10.3382/ps.2010-00978

Ball, R., Courtney-Martin, G., & Pencharz, P. (2006). The in vivo sparing of methionine by cysteine in sulfur amino acid requirements in animal models and adult humans. *Journal of Nutrition*, *136*, 1682S–1693S. https://doi.org/10.1093/jn/136.6.1682S

Banaszkiewicz, T. (2011). Nutritional value of soybean meal. In H. El-Shemy (Ed.), *Soybean and nutrition* (pp. 1–23). Intech Open. https://doi.org/10.5772/23306

Bandillo, N., Jarquin, D., Song, Q., Nelson, R., Cregan, P., Specht, J., & Lorenz, A. (2015). A population structure and genome-wide association analysis on the USDA soybean germplasm collection. *Plant Genome*, *8*(3), plantgenome2015.04.0024. https://doi.org/10.3835/plantgenome2015.04.0024

Bando, N., Tsuji, H., Yamanishi, R., Nio, N., & Ogawa, T. (1996). Identification of the glycosylation site of a major soybean allergen, Gly m Bd 30K. *Bioscience, Biotechnology, and Biochemistry*, *60*(2), 347–348. https://doi.org/10.1271/bbb.60.347

Bao, Y., Vuong, T., Meinhardt, C., Tiffin, P., Denny, R., Chen, S., Nguyen, H. T., Orf, J. H., & Young, N. D. (2014). Potential of association mapping and genomic selection to explore PI 88788 derived soybean cyst nematode resistance. *Plant Genome*, *7*(3), 1–13. https://doi.org/10.3835/plantgenome2013.11.0039

Beilinson, V., Chen, Z., Shoemaker, R., Fischer, R., Goldberg, R., & Nielsen, N. (2002). Genomic organization of glycinin genes in soybean. *Theoretical and Applied Genetics*, *104*, 1132–1140. https://doi.org/10.1007/s00122-002-0884-6

Bellaloui, N., Smith, J. R., Gillen, A. M., & Ray, J. D. (2010). Effect of maturity on seed sugars as measured on near-isogenic soybean (*Glycine max*) lines. *Crop Science*, *50*, 1978–1987. https://doi.org/10.2135/cropsci2009.10.0596

Bhat, J. A., Adeboye, K. A., Ganie, S. A., Barmukh, R., Hu, D., Varshney, R. K., & Yu, D. (2022). Genome-wide association study, haplotype analysis, and genomic prediction reveal the genetic basis of yield-related traits in soybean (Glycine max L.). *Frontiers in Genetics*, *13*, 953833. https://doi.org/10.3389/fgene.2022.953833

Bhusal, S. J., Orf, J., & Lorenz, A. J. (2022). Registration of M10-207102 soybean germplasm: A high-yielding, early-maturity line with elevated protein. *Journal of Plant Registration*, *16*, 132–136. https://doi.org/10.1002/plr2.20171

Bilyeu, K., Gillman, J. D., & LeRoy, A. R. (2011). Novel mutant allele combinations produce soybeans containing 1% linolenic acid in the seed oil. *Crop Science*, *51*, 259–264. https://doi.org/10.2135/cropsci2010.01.0044

Bilyeu, K., Palavalli, L., Sleper, D., & Beuselinck, P. (2005). Mutations in soybean microsomal omega-3 fatty acid desaturase genes reduce linolenic acid concentration in soybean seeds. *Crop Science*, *45*(5), 1830–1836. https://doi.og/10.2135/cropsci2004.0632

Bilyeu, K., Palavalli, L., Sleper, D., & Beuselinck, P. (2006). Molecular genetic resources for development of 1% linolenic acid soybeans. *Crop Science*, *46*(5), 1913–1918. https://doi.org/10.2135/cropsci2005.11-0426

Bilyeu, K., Ren, C., Nguyen, H. T., Herman, E., & Sleper, D. A. (2009). Association of a four-basepair insertion in the P34 gene with the low-allergen trait in soybean. *The Plant Genome*, *2*, 141–148. https://doi.org/10.3835/plantgenome2009.01.0006

Bilyeu, K., Škrabišová, M., Allen, D., Rajcan, I., Palmquist, D. E., Gillen, A., Mian, R., & Jo, H. (2018). The interaction of the soybean seed high oleic acid oil trait with other fatty acid modifications. *Journal of American Oil Chemical Society*, *95*(1), 39–49. https://doi/10.1002/aocs.12025

Bilyeu, K., & Wiebold, W. J. (2016). Environmental stability of seed carbohydrate profiles in soybeans containing different alleles of the raffinose synthase 2 (RS2) gene. *Journal of Agricultural and Food Chemistry*, *64*, 1071–1078. https://doi.org/10.1021/acs.jafc.5b04779

Bilyeu, K. D., Palavalli, L., Sleper, D. A., & Beuselinck, P. R. (2003). Three microsomal omega-3 fatty-acid desaturase genes contribute to soybean linolenic acid levels. *Crop Science*, *43*(5), 1833–1838. https://10.2135/cropsci2003.1833

Boehm, J., Jr., Nguyen, V., Tashiro, R., Anderson, D., Shi, C., Wu, X., Woodrow, L., Yu, K., Cui, Y., & Li, Z. (2018). Genetic mapping and validation of the loci controlling 7S α' and 11S A-type storage protein subunits in soybean [*Glycine max* (L.) Merr.]. *Theoretical and Applied Genetics*, *131*, 659–671. https://doi.org/10.1007/s00122-017-3027-9

Bolon, Y. T., Joseph, B., Cannon, S. B., Graham, M. A., Diers, B. W., Farmer, A. D., May, A. D., Muehlbauer, G. J., Specht, J. E., Zeng, J. T. W., & Vance, C. P. (2010). Complementary genetic and genomic approaches help characterize the linkage group I seed protein QTL in soybean. *BMC Plant Biology*, *10*, 41. https://doi.org/10.1186/1471-2229-10-41

Bosaz, L. B., Gerde, J. A., Borrás, L., Cipriotti, P. A., Ascheri, L., Campos, M., Gallo, S., & Rotundo, J. L. (2019). Management and environmental factors explaining soybean seed protein variability in central Argentina. *Field Crops Research*, *240*, 34–43. https://doi.org/10.1016/j.fcr.2019.05.007

Bradley, R. A., Atkinson, D., Hal1ser, H., Oldani, D., Green, J. P., & Stubbs, J. M. (1975). The structure, physical and chemical properties of the soybean glycinin. *Biochimica et Biophysica Acta*, *412*, 214–228. https://doi.org/10.1016/0005-2795(75)90036-7

Breene, W. M., Lin, S., Hardman, L., & Orf, J. (1988). Protein and oil content of soybeans from different geographic locations. *Journal of the American Oil Chemists' Society*, *65*, 1927–1931. https://doi.org/10.1007/BF02546009

Brummer, E. C., Graef, G. L., Orf, J., Wilcox, J. R., & Shoemaker, R. C. (1997). Mapping QTL for seed protein and oil content in eight soybean populations. *Crop Science*, *37*(2), 370–378. https://doi.org/10.2135/cropsci1997.0011183X003700020011x

Bubeck, D. M., Fehr, W. R., & Hammond, E. G. (1989). Inheritance of palmitic and stearic acid mutants of soybean. *Crop*

Science, *29*(3), 652–656. https://doi.org/10.2135/cropsci1989.0011183X002900030021x

Burton, J. W., Carter, T. E., Jr., & Wilson, R. F. (1999). Registration of 'Prolina' soybean. *Crop Science*, *39*, 294–295. https://doi.org/10.2135/cropsci1999.0011183X003900010066x

Burton, J. W., Miranda, L. M., Carter, T. E., Jr., & Bowman, D. T. (2012). Registration of 'NC-Miller' soybean with high yield and high seed-oil content. *Journal of Plant Registrations*, *6*, 294–297. https://doi.org/10.3198/jpr2012.01.0007crc

Byfield, G. E., Xue, H., & Upchurch, R. G. (2006). Two genes from soybean encoding soluble Δ9 stearoyl-ACP desaturases. *Crop Science*, *46*, 840–846. https://doi.org/10.2135/cropsci2005.06-0172

Cabrera-Bosquet, L., Crossa, J., von Zitzewitz, J., Serret, M. D., & Luis Araus, J. (2012). High-throughput phenotyping and genomic selection: The frontiers of crop breeding converge. *Journal of Integrative Plant Biology*, *54*(5), 312–320. https://doi.org/10.1111/j.1744-7909.2012.01116.x

Cabrera-Orozco, A., Jimenez-Martinez, C., & Davila-Ortiz, G. (2013). Soybean: Non-nutritional factors and their biological functionality. In H. El-Shemy (Ed.), *Soybean bio-active compounds* (pp. 387–410). IntechOpen. https://www.intechopen.com/chapters/39380

Cai, T., & Chang, K. C. (1999). Processing effect on soybean storage proteins and their relationship with tofu quality. *Journal of Agricultural and Food Chemistry*, *47*(2), 720–727. https://doi.org/10.1021/jf980571z

Candreva, A., Smaldini, P., Curciarello, R., Fossati, C., Docena, G., & Petruccelli, S. (2016). The major soybean allergen Gly m Bd 28K induces hypersensitivity reactions in mice sensitized to cow's milk proteins. *Journal of Agricultural Food Chemistry*, *64*(7), 1590–1599. https://doi.org/10.1021/acs/jafc.5b05623

Cao, L., Wang, Z., Ma, H., Liu, T., Ji, J., & Duan, K. (2022). Multiplex CRISPR/Cas9-mediated raffinose synthase gene editing reduces raffinose family oligosaccharides in soybean. *Frontiers in Plant Science*, *13*, 1048967. https://doi.org/10.3389/fpls.2022.1048967

Cao, Y., Li, S., Wang, Z., Chang, F., Kong, J., Gai, J., & Zhao, T. (2017). Identification of major quantitative trait loci for seed oil content in soybeans by combining linkage and genome-wide association mapping. *Frontiers in Plant Science*, *8*, 1222. https://doi.org/10.3389/fpls.2017.01222

Chang, S. K. C. (2015). Science and technology of tofu making. In Y. H. Hui & E. O. Evranuz (Eds.), *Handbook of fruits and vegetable processing* (2nd ed., pp. 817–840). CRC Press.

Chang, W. S., Lee, H. I., & Hungria, M. (2015). Soybean production in the Americas. In B. Lugtenberg (Ed.), *Principles of plant-microbe interactions: Microbes for sustainable agriculture* (pp. 393–400). Springer International Publishing.

Chappell, A. S., & Bilyeu, K. D. (2007). The low linolenic acid soybean line PI 361088B contains a novel GmFAD3A mutation. *Crop Science*, *47*(4), 1705–1710. https://doi.org/10.2135/cropsci2006.12.0783

Chaudhary, J., Patil, G. B., Sonah, H., Deshmukh, R. K., Vuong, T. D., Valliyodan, B., & Nguyen, H. T. (2015). Expanding omics resources for improvement of soybean seed composition traits. *Frontiers in Plant Science*, *6*, 1021. https://doi.org/10.3389/fpls.2015.01021

Chen, P., Huhn, M. R., Buss, G. R., Gunduz, I., & Wilson, J. H. (2001). *Interrelationship among seed quality traits of specialty soybeans for natto and tofu*. ASA-CSSA-SSSA International Annual Meetings.

Chen, P., Shannon, J. G., Ali, M. L., Scaboo, A., Crisel, M., Smothers, S., Clubb, M., Selves, S., Vieira, C. C., Mitchum, M. G., Nguyen, H. T., Li, Z., Bond, J., Meinhardt, C., Usovsky, M., Li, S., Mengistu, A., & Robbins, R. T. (2020). Registration of 'S14-9017GT' soybean cultivar with high yield, resistance to multiple diseases, and high seed oil content. *Journal of Plant Registrations*, *14*, 347–356. https://doi.org/10.1002/plr2.20011

Chen, P., Shannon, J. G., Lee, D., Granja, M. O., Vieira, C. C., Lee, Y. C., Ali, M. L., Nascimento, E. F., Scaboo, A., Crisel, M., Smothers, S., Clubb, M., Selves, S., Nguyen, H., Li, Z., Mitchum, M. G., Averitt, B., Bond, J., Meinhardt, C., … Moseley, D. (2023). Registration of 'S17-2243C': A non-genetically modified maturity group IV soybean cultivar with high yield and elevated oil concentration. *Journal of Plant Registrations*, *17*(2), 318–328. https://doi.org/10.1002/plr2.20276

Chen, P., Shannon, J. G., Scaboo, A., Crisel, M., Smothers, S., Clubb, M., Selves, S., Vieira, C. C., Ali, M. L., Lee, D., Nguyen, H. T., Li, Z., Mitchum, M. G., Bond, J., Meinhardt, C., Usovsky, M., Li, S., Mengistu, A., & Robbins, R. T. (2021). 'S13-1955C': A high-yielding conventional soybean with high oil content, multiple disease resistance, broad adaptation. *Journal of Plant Registrations*, *15*, 318–325. https://doi.org/10.1002/plr2.20112

Chen, P., Shannon, J. G., Scaboo, A., Crisel, M., Smothers, S., Clubb, M., Selves, S., Vieira, C. C., Ali, M. L., Lee, D., Nguyen, H. T., Li, Z., Mitchum, M. G., Bond, J. P., Meinhardt, C. G., Usovsky, M., Li, S., Mengistu, A., Zhang, B., … Robbins, R. T. (2022). Registration of 'S13-10592C', a high-yielding soybean cultivar with resistance to multiple diseases and elevated oil content. *Journal of Plant Registrations*, *16*, 252–261. https://doi.org/10.1002/plr2.20182

Chen, W., Song, K., Cai, Y., Li, W., Liu, B., & Liu, L. (2011). Genetic modification of soybean with a novel grafting technique: downregulating the FAD2-1 gene increases oleic acid content. *Plant Molecular Biology Reporter*, *29*(4), 866–874. https://doi.org/10.1007/s11105-011-0286-5

Chen, Z., Chen, Y., Xue, Z., Gao, X., Jia, Y., Wang, Y., Lu, Y., Zhang, J., Zhang, M., & Chen, H. (2020). Insight into the inactivation mechanism of soybean Bowman-Birk trypsin inhibitor (BBTI) induced by epigallocatechin gallate and epigallocatechin: Fluorescence, thermodynamics and docking studies. *Food Chemistry*, *303*, 125380. https://doi.org/10.1016/j.foodchem.2019.125380

Chen, Z., & Shoemaker, R. C. (1998). Four genes affecting seed traits in soybean map to linkage group F. *Journal of Heredity*, *89*, 211–215. https://doi.org/10.1093/jhered/89.3.211

Cho, T. J., Davies, C. S., & Nielsen, N. C. (1989). Inheritance and organization of glycinin genes in soybean. *Plant Cell*, *1*, 329–337. https://doi.org/10.1105/tpc.1.3.329

Choung, M.-G. (2010). Determination of sucrose content in soybean using near-infrared reflectance spectroscopy. *Journal of the Korean Society for Applied Biological Chemistry*, *53*(4), 478–484. https://doi.org/10.3839/jksabc.2010.073

Chung, J., Babka, H. L., Graef, G. L., Staswick, P. E., Lee, D. J., Cregan, P. B., Shoemaker, R. C., & Specht, J. E. (2003). The seed protein, oil, and yield QTL on soybean linkage group I. *Crop Science*, *43*, 1053–1067. https://doi.org/10.2135/cropsci2003.1053

Chung, J. I. (2009). A new soybean cultivar "Gaechuck # 1": Black soybean cultivar with lipoxygenase2, 3-free, Kunitz trypsin inhibitor-free and green cotyledon. *Korean Journal of Breeding Science*, *41*, 603–607. https://G704-000329.2009.41.4.026

Clemente, T. E., & Cahoon, E. B. (2009). Soybean oil: Genetic approaches for modification of functionality and total content. *Plant Physiology*, *151*(3), 1030–1040. https://doi.org/10.1104/pp.109.146282

Cober, E. R., & Voldeng, H. D. (2000). Developing high-protein, high-yield soybean populations and lines. *Crop Science*, *40*, 39–42. https://doi.org/10.2135/cropsci2000.40139x

Combs, R., & Bilyeu, K. (2019). Novel alleles of FAD2-1A induce high levels of oleic acid in soybean oil. *Molecular Breeding*, *39*(6), 79. https://doi.org/10.1007/s11032-019-0972-9

Coon, C. N., Leske, K. L., Akavanichan, O., & Cheng, T. K. (1990). Effect of oligosaccharide-free soybean meal on true metabolizable energy and fiber digestion in adult roosters. *Poultry Science*, *69*(5), 787–793. https://doi.org/10.3382/ps.0690787

Cregan, P. B., Jarvik, T., Bush, A. L., Shoemaker, R. C., Lark, K. G., Kahler, A. L., Kaya, N., Vantoai, T. T., Lohnes, D. G., Chung, J., & Specht, J. E. (1999). An integrated genetic linkage map of the soybean genome. *Crop Science*, *20*, 1464–1490. https://doi.org/10.2135/cropsci1999.3951464x

Cromwell, G. L. (1999). Soybean meal—The “gold standard”. *The Farmer’s Pride, KPPA News*, *11*(20), 00.

Crossa, J., Pérez-Rodríguez, P., Cuevas, J., Montesinos-López, O., Jarquín, D., de los Campos, G., Burgueño, J., González-Camacho, J. M., Pérez-Elizalde, S., Beyene, Y., Dreisigacker, S., Singh, R., Zhang, X., Gowda, M., Roorkiwal, M., Rutkoski, J., & Varshney, R. K. (2017). Genomic selection in plant breeding: Methods, models, and perspectives. *Trends in Plant Science*, *22*(11), 961–975. https://doi.org/10.1016/j.tplants.2017.08.011

Cunicelli, M. J., Bhandari, H. S., Chen, P., Sams, C. E., Mian, M. A. R., Mozzoni, L. A., Smallwood, C. J., & Pantalone, V. R. (2019). Effect of a mutant Danbaekkong allele on soybean seed yield, protein, and oil concentration. *Journal of the American Oil Chemists’ Society*, *96*(8), 927–935. https://doi.org/10.1002/aocs.12261

Danaei, G., Ding, E. L., Mozaffarian, D., Taylor, B., Rehm, J., Murray, C. J. L., & Ezzati, M. (2009). The preventable causes of death in the United States: Comparative risk assessment of dietary, lifestyle, and metabolic risk factors. *PLoS Medicine*, *6*(4), e1000058. https://doi.org/10.1371/journal.pmed.1000058

De Maria, M., Robinson, E. J., Kangile, J. R., Kadigi, R., Dreoni, I., Couto, M., Howai, N., & Peci, J. (2020). *Global soybean trade—The geopolitics of a bean*. UK Research and Innovation Global Challenges Research Fund (UKRI GCRF), Trade, Development, and the Environment Hub. https://doi.org/10.34892/7yn1-k494

Demorest, Z. L., Coffman, A., Baltes, N. J., Stoddard, T. J., Clasen, B. M., Luo, S., Retterath, A., Yabandith, A., Gamo, M. E., Bissen, J., Mathis, L., Voytas, D. F., & Zhang, F. (2016). Direct stacking of sequence-specific nuclease-induced mutations to produce high oleic and low linolenic soybean oil. *BMC Plant Biology*, *16*(1), 225. https://doi.org/10.1186/s12870-016-0906-1

de Azevedo, L., Peixoto, Moellers, T. C., Zhang, J., Lorenz, A. J., Bhering, L. L., Beavis, W. D., & Singh, A. K. (2017). Leveraging genomic prediction to scan germplasm collection for crop improvement. *PLoS One*, *12*(6), e0179191. https://doi.org/10.1371/journal.pone.0179191

de Souza, R. J., Mente, A., Maroleanu, A., Cozma, A. I., Ha, V., Kishibe, T., Uleryk, E., Budylowski, P., Schunemann, H., Beyene, J., & Ananda, S. S. (2015). Intake of saturated and trans unsaturated fatty acids and risk of all cause mortality, cardiovascular disease, and type 2 diabetes: Systematic review and meta-analysis of observational studies. *BMJ*, *351*, h3978. https://doi.org/10.1136/bmj.h3978

Dierking, E. C., & Bilyeu, K. D. (2008). Association of a soybean raffinose synthase gene with low raffinose and stachyose seed phenotype. *The Plant Genome*, *1*(2), 135–145. https://doi.org/10.3835/plantgenome2008.06.0321

Dierking, E. C., & Bilyeu, K. D. (2009). New sources of soybean seed meal and oil composition traits identified through TILLING. *BMC Plant Biology*, *9*(1), 89. https://doi.org/10.1186/1471-2229-9-89

Diers, B. W., Beilinson, V., Nielsen, N. C., & Shoemaker, R. C. (1994). Genetic mapping of the Gy4 and Gy5 glycinin genes in soybean and analysis of a variant of Gy4. *Theoretical and Applied Genetic*, *89*, 297–304. https://doi.org/10.1007/BF00225158

Diers, B. W., Keim, P., Fehr, W. R., & Shoemaker, R. C. (1992). RFLP analysis of soybean seed protein and oil content. *Theoretical and Applied Genetics*, *83*, 608–612. https://doi.org/10.1007/BF00226905

Diers, B. W., & Shoemaker, R. C. (1992). Restriction fragment length polymorphism analysis of soybean fatty acid content. *Journal of American Oil Chemists’ Society*, *69*(12), 1242–1244. https://doi.org/10.1007/BF02637690

Do, P. T., Nguyen, C. X., Bui, H. T., Tran, L. T. N., Stacey, G., Gillman, J. D., Zhang, Z. J., & Stacey, M. G. (2019). Demonstration of highly efficient dual gRNA CRISPR/Cas9 editing of the homeologous GmFAD2–1A and GmFAD2–1B genes to yield a high oleic, low linoleic and α-linolenic acid phenotype in soybean. *BMC Plant Biology*, *19*(1), 311. https://doi.org/10.1186/s12870-019-1906-8

Dozier, W. A., III, & Hess, J. B. (2011). Soybean meal quality and analytical techniques. In H. El-Shemy (Ed.), *Soybean and nutrition* (pp. 111–124). IntechOpen. https://www.intechopen.com/chapters/19977

Duhnen, A., Gras, A., Teyssèdre, S., Romestant, M., Claustres, B., Daydé, J., & Mangin, B. (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. https://doi.org/10.2135/cropsci2016.06.0496

Elayabharathy, V. (2020). *Reducing the negative impacts of phytate in feed*. Bentoli. https://www.bentoli.com/reducing-the-negative-impacts-of-phytate/

Fallen, B., Cunicelli, M., Wyman, C., & Pantalone, V. (2022). Registration of TN13-4304 soybean germplasm with good yield, high meal protein, and resistance to peanut and southern root knot nematode. *Journal of Plant Registrations*, *16*, 423–429. https://doi.org/10.1002/plr2.20191

Fallen, B. D., Hatcher, C. N., Allen, F. L., Kopsell, D. A., Saxton, A. M., Chen, P., Kantartzi, S. K., Cregan, P. B., Hyten, D. L., & Pantalone, V. R. (2013). Soybean seed amino acid content QTL detected using the universal soy linkage panel 1.0 with 1,536 SNPs. *Journal of Plant Genome Sciences*, *1*, 68–79. https://doi.org/10.5147/jpgs.2013.0089

Fehr, W. R. (2007). Breeding for modified fatty acid composition in soybean. *Crop Science*, *47*, 72–87. https://doi.org/10.2135/cropsci2007.04.0004IPBS

Fehr, W. R., Caviness, C. E., Burmood, D. T., & Pennington, J. S. (1971). Stage of development descriptions for soybeans, *Glycine max* (L.) Merrill. *Crop Science*, *11*, 929–931. https://doi.org/10.2135/cropsci1971.0011183X001100060051x

Feng, X., Yang, S., Tang, K., Zhang, Y., Leng, J., Ma, J., Wang, Q., & Feng, X. (2019). GmPGL1, a thiamine thiazole synthase, is required for the biosynthesis of thiamine in soybean. *Frontiers in Plant Science*, *10*, 1546. https://doi.org/10.3389/fpls.2019.01546

Ficht, A., Bruce, R., Torkamaneh, D., Grainger, C. M., Eskandari, M., & Rajcan, I. (2022). Genetic analysis of sucrose concentration in soybean seeds using a historical soybean genomic panel. *Theoretical and Applied Genetics*, *135*, 1375–1383. https://doi.org/10.1007/s00122-022-04040-z

Fischer, R. L., & Goldberg, R. B. (1982). Structure and flanking regions of soybean seed protein genes. *Cell*, *29*, 651–660. https://doi.org/10.1016/0092-8674(82)90181-7

Flavin, D. F. (1982). The effects of soybean trypsin inhibitors on the pancreas of animals and man: A review. *Veterinary and Human Toxicology*, *24*, 25–28. https://europepmc.org/article/med/7036513

Fliege, C. E., Ward, R. A., Vogel, P., Nguyen, H., Quach, T., Guo, M., Gomes Viana, J. P., dos Santos, L. B., Specht, J. E., Clemente, T. E., Hudson, M. E., & Diers, B. W. (2022). Fine mapping and cloning of the major seed protein quantitative trait loci on soybean chromosome 20. *The Plant Journal*, *110*(1), 114–128. https://doi.org/10.1111/tpj.15658

Food and Agriculture Organization of the United Nations (FAO). (2023). *FAOSTAT*. https://www.fao.org/faostat/en/#data/QCL

Friedman, M., & Brandon, D. L. (2001). Nutritional and health benefits of soy proteins. *Journal of Agriculture and Food Chemistry*, *49*, 1069–1086. https://doi.org/10.1021/jf0009246

Gale, F., Valdes, C., & Ash, M. (2019). *Interdependence of China, United States, and Brazil in soybean trade* (Oil Crops Outlook No. OCS-19F-01). Economic Research Service, U.S. Department of Agriculture.

Gao, Y., Biyashev, R. M., Maroof, M. A. S., Glover, N. M., Tucker, D. M., & Buss, G. R. (2008). Validation of low-phytate QTL and evaluation of seedling emergence of low-phytate soybeans. *Crop Science*, *48*, 1355–1364. https://doi.org/10.2135/cropsci2007.11.0633

Gillman, J. D., Pantalone, V. R., & Bilyeu, K. (2009). The low phytic acid phenotype in soybean line CX1834 is due to mutations in two homologs of the maize low phytic acid gene. *The Plant Genome*, *2*, 179–190. https://doi.org/10.3835/plantgenome2008.03.0013

Goettel, W., Zhang, H., Li, Y., Qiao, Z., Jiang, H., Hou, D., Song, Q., Pantalone, V. R., Song, B.-H., Yu, D., & An, Y.-Q. C. (2022). POWR1 is a domestication gene pleiotropically regulating seed quality and yield in soybean. *Nature Communications*, *13*, 3051. https://doi.org/10.1038/s41467-022-30314-7

Goßner, S., Yuan, F., Zhou, C., Tan, Y., Shu, Q., & Engel, K. H. (2019). Impact of cross-breeding of low phytic acid MIPS1 and IPK1 soybean (*Glycine max* L. Merr.) mutants on their contents of inositol phosphate isomers. *Journal of Agricultural and Food Chemistry*, *67*(1), 247–257. https://doi.org/10.1021/acs.jafc.8b06117

Graef, G. L., Fehr, W. R., & Hammond, E. G. (1985). Inheritance of three stearic acid mutants of soybean. *Crop Science*, *25*(6), 1076–1079. https://doi.org/10.2135/cropsci1985.0011183X002500060043x

Graef, G. L., Fehr, W. R., Miller, L. A., Hammond, E. G., & Cianzo, S. R. (1988). Inheritance of fatty acid composition in a soybean mutant with low linolenic acid. *Crop Science*, *28*(1), 55–58. https://doi.org/10.2135/cropsci1988.0011183X002800010013x

Grant, D., Nelson, R. T., Cannon, S. B., & Shoemaker, R. C. (2010). SoyBase, the USDA-ARS soybean genetics and genomics database. *Nucleic Acids Research*, *38*, D843–D846. https://doi.org/10.1093/nar/gkp798

Grieshop, C. M., & Fahey, G. C. (2001). Comparison of quality characteristics of soybean from Brazil, China, and the United States. *Journal of Agricultural and Food Chemistry*, *49*(5), 2669–2673. https://doi.org/10.1021/jf0014009

Guan, R.-X., Fang, X. Q., Chang, R. Z., & Qiu, L. J. (2012). Hypoallergenic soybean, from genes to cultivar. In Z.-S. Gao, M. Zheng, L. J. W. J. Gilissen, H.-H. Shen, & L. J. Frewer (Eds.), *Multidisciplinary approaches to allergies* (pp. 347–358). Springer. https://doi.org/10.1007/978-3-642-31609-8_21

Guinn, J. M. (2002). *Domestic quality standards and trading rules and recommended export contract specifications for U.S. soybeans and products*. U.S. Soybean Export Council. http://ussec.org/wp-content/uploads/2012/08/Guinn_Quality_Standards_Trading_Rules2002.pdf

Guo, J., & Yang, X. Q. (2015). Texture modification of soy-based products. In J. Chen & A. Rosenthal (Eds.), *Modifying food texture* (pp. 237–255). Woodhead Publishing Series. https://doi.org/10.1016/B978-1-78242-333-1.00011-5

Hagely, K. B., Jo, H., Kim, J. H., Hudson, K. A., & Bilyeu, K. (2020). Molecular-assisted breeding for improved carbohydrate profiles in soybean seed. *Theoretical and Applied Genetics*, *133*, 1189–1200. https://doi.org/10.1007/s00122-020-03541-z

Hagely, K. B., Palmquist, D., & Bilyeu, K. D. (2013). Classification of distinct seed carbohydrate profiles in soybean. *Journal of Agricultural and Food Chemistry*, *61*(5), 1105–1111. https://doi.org/10.1021/jf303985q

Ham, G. E., Liener, I. E., Evans, S. D., Frazier, R. D., & Nelson, W. W. (1975). Yield and component of soybean seed as affected by N and S fertilization. *Agronomy Journal*, *67*(3), 293–297. https://doi.org/10.2134/agronj1975.00021962006700030004x

Han, F. X., Ding, A. L., & Sun, J. M. (2002). Development of a new soybean variety with null trypsin inhibitor and lipoxygenase 2.3 genes–zhonghuang 16 and its cultivation practices. *Acta Genetica Sinica*, *29*, 1105–1110.

Han, J., Guo, B., Guo, Y., Zhang, B., Wang, X., & Qiu, L.-J. (2019). Creation of early flowering germplasm of soybean by CRISPR/Cas9 technology. *Frontiers in Plant Science*, *10*, 1446. https://www.frontiersin.org/articles/10.3389/fpls.2019.01446

Han, Y., Li, D., Zhao, G., Zhao, X., Jiang, Z., Hu, H., Wu, L., Wang, Y., Gao, Y., Li, Y., Zeng, G., Meng, F., Teng, W., & Li, W. (2016). Dynamic quantitative trait loci underlies isoflavone accumulation in soybean seed. *Plant Breeding*, *135*(3), 335–341. https://doi.org/10.1111/pbr.12354

Hannah, R., & Max, R. (2021). Forests and deforestation. https://ourworldindata.org/forests-and-deforestation

Harada, J. J., Barker, S. J., & Goldberg, R. B. (1989). Soybean β-conglycinin genes are clustered in several DNA regions and are regulated by transcriptional and posttranscriptional processes. *Plant Cell*, *1*, 415–425. https://doi.org/10.1105/tpc.1.4.415

Hartman, G. L., West, E. D., & Herman, T. K. (2011). Crops that feed the World 2. Soybean—Worldwide production, use, and constraints caused by pathogens and pests. *Food Security*, *3*, 5–17. https://doi.org/10.1007/s12571-010-0108-x

Hayashi, M., Kitamura, K., & Harada, K. (2009). Genetic mapping of Cgdef gene controlling accumulation of 7S globulin (b-conglycinin) subunits in soybean seeds. *Journal of Heredity*, *100*, 802–806. https://doi.org/10.1093/jhered/esp046

Hemingway, J., Schnebly, S. R., & Rajcan, I. (2021). Accuracy of genomic prediction for seed oil concentration in high-oleic soybean populations using a low-density marker panel. *Crop Science*, *61*(6), 4012–4021. https://doi.org/10.1002/csc2.20607

Herman, E. M., Helm, R. M., Jung, R., & Kinney, A. J. (2003). Genetic modification removes an immunodominant allergen from soybean. *Plant Physiology*, *132*(1), 36–43. https://doi.org/10.1104/pp.103.021865

Hill, G. D. (2003). Plant antinutritional factors: Characteristics. In B. Caballero (Ed.), *Encyclopedia of food sciences and nutrition* (2nd ed., pp. 4578–4587). Academic Press.

Hitz, W. D., Carlson, T. J., Kerr, P. S., & Sebastian, A. (2002). Biochemical and molecular characterization of a mutation that confers a decreased raffinosaccharide and phytic acid phenotype on soybean seeds. *Plant Physiology*, *128*, 650–660. https://doi.org/10.1104/pp.010585

Hou, A., Chen, P., Alloatti, J., Li, D., Mozzoni, L., Zhang, B., & Shi, A. (2009). Genetic variability of seed sugar content in worldwide soybean germplasm collections. *Crop Science*, *49*(3), 903–912. https://doi.org/10.2135/cropsci2008.05.0256

Howell, R. W., & Cartter, J. L. (1958). Physiological factors affecting composition of soybeans: II. response of oil and other constituents of soybeans to temperature under controlled conditions. *Agronomy Journal*, *50*, 664–667. https://doi.org/10.2134/agronj1958.00021962005000110007x

Hsu, S. H., Hadley, H. H., & Hymowitz, T. (1973). Changes in carbohydrate contents of germinating soybean seeds. *Crops Science*, *13*(4), 407–410. https://doi.org/10.2135/cropsci1973.0011183X001300040004x

Hudson, K., & Hudson, M. (2021). Genetic variation for seed oil biosynthesis in soybean. *Plant Molecular Biology Reporter*, *39*, 700–709. https://doi.org/10.1007/s11105-020-01276-1

Hurburgh, C. R., Jr., Brumm, T. J., Guinn, J. M., & Hartwig, R. A. (1990). Protein and oil patterns in U. S. and world soybean markets. *Journal of the American Oil Chemists' Society*, *67*(12), 966–973. https://doi.org/10.1007/BF02541859

Hurrell, R. F. (2003). Influence of vegetable protein sources on trace element and mineral bioavailability. *The Journal of Nutrition*, *133*(9), 2973S–2977S. https://doi.org/10.1093/jn/133.9.2973S

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

Hymowitz, T. (1986). Genetics and breeding of soybeans lacking the Kunitz trypsin inhibitor. In M. Friedman (Ed.), *Nutritional and toxicological significance of enzyme inhibitors in foods* (pp. 291–298). Springer US.

Hymowitz, T., & Harlan, J. R. (1983). Introduction of soybean to North America by Samuel Bowen in 1765. *Economic Botany*, *37*, 371–379. https://doi.org/10.1007/BF02904196

Hyten, D. L., Pantalone, V. R., Sams, C. E., Saxton, A. M., Landau-Ellis, D., Stefaniak, T., & Schmidt, M. E. (2004). Seed quality QTL in a prominent soybean population. *Theoretical and Applied Genetics*, *109*(3), 552–561. https://doi.org/10.1007/s00122-004-1661-5

Hyten, D. L., Pantalone, V. R., Saxton, A. M., Schmidt, M. E., & Sams, C. E. (2004). Molecular mapping and identification of soybean fatty acid modifier quantitative trait loci. *Journal of American Oil Chemists' Society*, *81*(12), 1115–1118. https://doi.org/10.1007/s11746-004-1027-z

Ibáñez, M. A., de Blas, C., Cámara, L., & Mateos, G. G. (2020). Chemical composition, protein quality and nutritive value of commercial soybean meals produced from beans from different countries: A meta-analytic study. *Animal Feed Science and Technology*, *267*, 114531. https://doi.org/10.1016/j.anifeedsci.2020.114531

Institute of Medicine. (2001). *Dietary reference intakes for vitamin A, vitamin K, arsenic, boron, chromium, copper, iodine, iron, manganese, molybdenum, nickel, silicon, vanadium, and zinc*. National Academies Press. http://www.ncbi.nlm.nih.gov/books/NBK222310/

Jaffe, G. (1981). Phytic acid in soybeans. *Journal of the American Oil Chemists' Society*, *58*, 493–495. https://doi.org/10.1007/BF02582411

James, A. T., & Yang, A. (2016). Interactions of protein content and globulin subunit composition of soybean proteins in relation to tofu gel properties. *Food Chemistry*, *194*, 284–289. https://doi.org/10.1016/j.foodchem.2015.08.021

Jarquín, D., Crossa, J., Lacaze, X., Du Cheyron, P., Daucourt, J., Lorgeou, J., Piraux, F., Guerreiro, L., Pérez, P., Calus, M., Burgueño, J., & de los Campos, G. (2014). A reaction norm model for genomic selection using high-dimensional genomic and environmental data. *Theoretical and Applied Genetics*, *127*(3), 595–607. https://doi.org/10.1007/s00122-013-2243-1

Jarquín, D., Kocak, K., Posadas, L., Hyma, K., Jedlicka, J., Graef, G., & Lorenz, A. (2014). Genotyping by sequencing for genomic prediction in a soybean breeding population. *BMC Genomics*, *15*(1), 740. https://doi.org/10.1186/1471-2164-15-740

Jarquin, D., Specht, J., & Lorenz, A. (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. https://doi.org/10.1534/g3.116.031443

Jeong, J. E., Kulkarni, K., Chang, J. H., Ha, B. K., Kang, S. T., Bilyeu, K., Jo, H., Song, J. T., & Lee, J. D. (2018). A novel allele of GmSACPD-C associated with high seed stearic acid concentration in an EMS-induced mutant PE980 in soybean. *Crop Science*, *58*, 192–203. https://doi.org/10.2135/cropsci2017.05.0313

Jeong, K. H., Choi, M. S., Lee, S. K., Seo, M. J., Hwang, T. Y., Yun, H. T., Kim, H. S., Kim, J. T., Kwon, Y. U., & Kim, Y. H. (2013). Development of low Gly m Bd 30K (P34) allergen breeding lines using molecular marker in soybean [*Glycine max* (L.) Merr.]. *Plant Breeding and Biotechnology*, *1*, 298–306. https://doi.org/10.9787/PBB.2013.1.3.298

Jia, Y., & Jannink, J.-L. (2012). Multiple-trait genomic selection methods increase genetic value prediction accuracy. *Genetics*, *192*(4), 1513–1522. https://doi.org/10.1534/genetics.112.144246

Jing, Y., GuoJie, X., Qian, D., Li, S., DongQuan, G., Lu, N., & XiangDong, Y. (2016). Effects of different soybean genotypes on the transformation efficiency of soybean and analysis of the T-DNA insertions in the soybean genome. *Soybean Science*, *35*(4), 562–567. https://www.cabdirect.org/cabdirect/abstract/20163374088

Jo, H., Lee, J. D., & Bilyeu, K. D. (2018). Environmental stability of carbohydrate profiles in different soybean genotypes. *Crop Science*, *58*(2), 773–782. https://doi.org/10.2135/cropsci2017.08.0497

Jo, H., Lorenz, A. J., Rainey, K. M., Shannon, J. G., Chen, P., & Bilyeu, K. D. (2019). Environmental stability study of soybeans with modified carbohydrate profiles in maturity groups 0 to v. *Crop Science*, *59*(4), 1531–1543. https://doi.org/10.2135/cropsci2018.09.0600

Jofuko, J., & Goldberg, R. B. (1989). Kunitz trypsin inhibitor genes are differentially expressed during the soybean life cycle and in transformed tobacco plants. *The Plant Cell*, *1*(11), 1079–1093. https://doi.org/10.2307/3869025

John, P. G. K. (2008). Sugar syrup: The new energy feed for poultry. *World Poultry Science*, *24*, 12–13.

Jun, T. H., Van, K., Kim, M. Y., Lee, S. H., & Walker, D. R. (2008). Association analysis using SSR markers to find QTL for seed protein content in soybean. *Euphytica*, *162*(2), 179–191. https://doi.org/10.1007/s10681-007-9491-6

Kang, I. J., Matsumura, Y., & Mori, T. (1991). Characteristics of texture and mechanical properties of heat-induced soy protein gels. *Journal of the American Oil Chemists' Society*, *68*, 339–345. https://doi.org/10.1007/BF02657690

Karl-Lilienthal, L. K., Grieshop, C. M., Merchen, N. R., Mahan, D. C., & Fahey, G. C., Jr. (2004). Chemical composition and protein quality comparisons of soybeans and soybean meals from five leading soybean-producing countries. *Journal of Agricultural and Food Chemistry*, *52*, 6193–6199. https://doi.org/10.1021/jf049795

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

Kerr, P. S., & Sebastian, S. A. (2000). *Soybean products with improved carbohydrate composition and soybean plants* (U.S. Patent No. US6147193A). U.S. Patent and Trademark Office.

Khandaker, L., Akond, M., Liu, S., Kantartzi, S. K., Meksem, K., Bellaloui, N., Lightfoot, D. A., & Kassem, M. A. (2015). Mapping of QTL associated with seed amino acids content in "MD96-5722" by "Spencer" RIL population of soybean using SNP markers. *Food and Nutrition Sciences*, *6*(11), 974–984. https://doi.org/10.4236/fns.2015.611101

Kim, H. K., Kang, S. T., Cho, J. H., Choung, M. G., & Suh, D. Y. (2005). Quantitative trait loci associated with oligosaccharide and sucrose contents in soybean (*Glycine max* L.). *Journal of Plant Biology*, *48*, 106–112. https://doi.org/10.1007/BF03030569

Kim, H. K., Kang, S. T., & Oh, K. W. (2006). Mapping of putative quantitative trait loci controlling the total oligosaccharide and sucrose content of *Glycine max* seeds. *Journal of Plant Research*, *119*, 533–538. https://doi.org/10.1007/s10265-006-0004-9

Kim, M. S., Park, M. J., Jeong, W. H., Nam, K. C., & Chung, J. I. (2006). SSR marker tightly linked to the Ti locus in Soybean [*Glycine max* (L.) Merr.]. *Euphytica*, *152*, 361–366. https://doi.org/10.1007/s10681-006-9223-3

Kim, Y., & Wicker, L. (2005). Soybean cultivars impact quality and function of soymilk and tofu. *Journal of the Science of Food and Agriculture*, *85*, 2514–2518. https://doi.org/10.1002/jsfa.2287

Kitamura, K. (1984). Biochemical characterization of lipoxygenase lacking mutants, Ll-less, L-2-less, and L-3-less soybeans. *Agricultural and Biological Chemistry*, *48*, 2339–2346. https://doi.org/10.1080/00021369.1984.10866488

Koo, S. C., Seo, J. S., Park, M. J., Cho, H. M., Park, M. S., Choi, C. W., Jung, W. H., Lee, K. H., Jin, B. J., Kim, S. H., Shim, S. I., Chung, J. S., Chung, J. I., & Kim, M. C. (2013). Identification of molecular mechanism controlling P34 gene expression in soybean. *Plant Biotechnology Reports*, *7*, 331–338. https://doi.org/10.1007/s11816-012-0267-7

Krishnan, H. B., & Jez, J. M. (2018). Review: The promise and limits for enhancing sulfur-containing amino acid content of soybean seed. *Plant Science*, *272*, 14–21. https://doi.org/10.1016/j.plantsci.2018.03.030

Kumar, V., Rani, A., Goyal, L., Dixit, A. K., Manjaya, J. G., Dev, J., & Swamy, M. (2010). Sucrose and raffinose family oligosaccharides (RFOs) in soybean seeds as influenced by genotype and growing location. *Journal of Agriculture and Food Chemistry*, *58*, 5081–5085. https://doi.org/10.1021/jf903141s

Kumar, V., Rani, A., & Mourya, V. (2011). Marker-assisted selection for development of Kunitz trypsin inhibitor free soybean varieties: I: Parental polymorphism survey using SSR markers. *The Indian Society of Genetics and Plant Breeding*, *71*, 372–376.

Kumar, V., Rani, A., & Rawal, R. (2013). Deployment of gene specific marker in development of Kunitz trypsin inhibitor free soybean genotype. *NISCAIR-CSIR*, *51*, 1125–1129.

Kumar, V., Rani, A., Solanki, S., & Hussain, S. M. (2006). Influence of growing environment on the biochemical composition and physical characteristics of soybean seed. *Journal of Food Composition and Analysis*, *19*(2–3), 188–195. https://doi.org/10.1016/j.jfca.2005.06.005

Lakhssassi, N., Zhou, Z., Liu, S., Piya, S., Cullen, M. A., Baze, A., Knizia, D., Patil, G. B., Badad, O., Embaby, M. G., Meksem, J., Lakhssassi, A., AbuGhazaleh, A., Hewezi, T., & Meksem, K. (2020). Soybean tilling-by-sequencing+ reveals the role of novel GmSACPD members in unsaturated fatty acid biosynthesis while maintaining healthy nodules. *Journal of Experimental Botany*, *71*(22), 6969–6987. https://doi.org/10.1093/jxb/eraa402

Le, H., Nguyen, N. H., Ta, D. T., Le, T. N. T., Bui, T. P., Le, N. T., Nguyen, C. X., Rolletschek, H., Stacey, G., Stacey, M. G., Phem, N. B., Do, P. T., & Chu, H. H. (2020). CRISPR/Cas9-mediated knockout of galactinol synthase-encoding genes reduces raffinose family oligosaccharide levels in soybean seeds. *Frontiers in Plant Science*, *11*, 612942. https://doi.org/10.3389/fpls.2020.612942

Lee, J. D., Shannon, G., & Choung, M. G. (2010). Selection for protein content in soybean from single F2 seed by near infrared reflectance spectroscopy. *Euphytica*, *172*, 117–123. https://doi.org/10.1007/s10681-009-0067-5

Lee, K. J., Hwang, J. E., Velusamy, V., Ha, B. K., Kim, J. B., Kim, S. H., Ahn, J. W., Kang, S. Y., & Kim, D. S. (2014). Selection and molecular characterization of a lipoxygenase-free soybean mutant line induced by gamma irradiation. *Theoretical Applied Genetics*, *127*, 2405–2413. https://doi.org/10.1007/s00122-014-2385-9

Lee, S., Van, K., Sung, M., Nelson, R., LaMantia, J., McHale, L. K., & Mian, M. A. (2019). Genome-wide association study of seed protein, oil and amino acid contents in soybean from maturity groups I to IV. *Theoretical and Applied Genetics*, *132*(6), 1639–1659. https://doi.org/10.1007/s00122-019-03304-5

Lee, S. H., Bailey, M. A., Mian, M. A. R., Carter, T. E., Shipe, E. R., Ashley, D. A., Parrott, W. A., Hussey, R. S., & Boerma, H. R. (1996). RFLP loci associated with soybean seed protein and oil content across populations and locations. *Theoretical and Applied Genetics*, *93*(5–6), 649–657. https://doi.org/10.1007/BF00224058

Lei, X., Ku, P. K., Miller, E. R., Ullrey, D. E., & Yokoyama, M. T. (1993). Supplemental microbial phytase improves bioavailability of dietary zinc to weanling pigs. *The Journal of Nutrition*, *123*(6), 1117–1123. https://doi.org/10.1093/jn/123.6.1117

Lei, X. G., & Porres, J. M. (2011). Phytase: An enzyme to improve soybean nutrition. In H. El-Shemy (Ed.), *Soybean and nutrition* (pp. 67–80). IntechOpen. https://doi.org/10.5772/20128

Lenis, J. M., Gillman, J. D., Lee, J. D., Shannon, J. G., & Bilyeu, K. D. (2010). Soybean seed lipoxygenase genes: Molecular characterization and development of molecular marker assays. *Theoretical Applied Genetics*, *120*, 1139–1149. https://doi.org/10.1007/s00122-009-1241-9

Li, C., & Zhang, Y. M. (2011). Molecular evolution of glycinin and β-conglycinin gene families in soybean (*Glycine max* L. Merr.). *Heredity*, *106*, 633–641. https://doi.org/10.1038/hdy.2010.97

Li, H., Liu, H., Han, Y., Wu, X., Teng, W., Liu, G., & Li, W. (2010). Identification of QTL underlying vitamin E contents in soybean seed among multiple environments. *Theoretical and Applied Genetics*, *120*(7), 1405–1413. https://doi.org/10.1007/s00122-010-1264-2

Li, L., Zheng, W., Zhu, Y., Ye, H., Tang, B., Arendsee, Z. W., Jones, D., Li, R., Ortiz, D., Zhao, X., Du, C., Nettleton, D., Scott, M. P., Fernandez, M. G. S., Yin, Y., & Wurtele, E. S. (2015). QQS orphan gene regulates carbon and nitrogen partitioning across species via NF-YC interactions. *Proceedings of the National Academy of Sciences*, *112*(47), 14734–14739. https://doi.org/10.1073/pnas.1514670112

Li, M., Li, X., Zhou, Z., Wu, P., Fang, M., Pan, X., Lin, Q., Luo, W., Wu, G., & Li, H. (2016). Reassessment of the four yield-related genes Gn1a, DEP1, GS3, and IPA1 in rice using a CRISPR/Cas9 system. *Frontiers in Plant Science*, *7*, 377. https://doi.org/10.3389/fpls.2016.00377

Liener, I. E. (1994). Implications of antinutritional components in soybean foods. *Critical Reviews in Food Science and Nutrition*, *34*, 31–67. https://doi.org/10.1080/10408399409527649

Lin, W., Kuang, H., Bai, M., Jiang, X., Zhou, P., Li, Y., Chen, B., Huarong, L., & Guan, Y. (2023). Multiplex genome editing targeting soybean with ultra-low anti-nutritive oligosaccharides. *The Crop Journal*, *11*(3), 825–831. https://doi.org/10.1016/j.cj.2023.01.001

Liu, S., Zhang, M., Feng, F., & Tian, Z. (2020). Toward a "green revolution" for soybean. *Molecular Plant*, *13*(5), 688–697. https://doi.org/10.1016/j.molp.2020.03.002

Liu, T., Ji, J., Cheng, Y., Zhang, S., Wang, Z., Duan, K., & Wang, Y. (2023). CRISPR/Cas9-mediated editing of GmTAP1 confers enhanced resistance to *Phytophthora sojae* in soybean. *Journal of Integrative Plant Biology*, *65*(7), 1609–1612. https://doi.org/10.1111/jipb.13476

Ma, X., Sun, P., He, P., Han, P., Wang, J., Qiao, S., & Li, D. (2010). Development of monoclonal antibodies and a competitive ELISA detection method for glycinin, an allergen in soybean. *Food Chemistry*, *121*, 546–551. https://doi.org/10.1016/j.foodchem.2009.12.045

Ma, Y., Kan, G., Zhang, X., Wang, Z., Zhang, W., Du, H., & Yu, D. (2016). Quantitative trait loci (QTL) mapping for glycinin and β conglycinin contents in soybean (*Glycine max* L. Merr.). *Journal of Agricultural and Food Chemistry*, *64*, 3473–3483. https://doi.org/10.1021/acs.jafc.6b00167

Maga, J. A. (1982). Phytate: Its chemistry, occurrence, food interactions, nutritional significance, and methods of analysis. *Journal of Agricultural and Food Chemistry*, *30*(1), 1–9. https://doi.org/10.1021/jf00109a001

Malle, S., Eskandari, M., Morrison, M., & Belzile, F. (2020). Genome-wide association identifies several QTL controlling cysteine and methionine content in soybean seed including some promising candidate genes. *Scientific Reports*, *10*, 21812. https://doi.org/10.1038/s41598-020-78907-w

Malle, S., Morrison, M., & Belzile, F. (2020). Identification of loci controlling mineral element concentration in soybean seeds. *BMC Plant Biology*, *20*, 419. https://doi.org/10.1186/s12870-020-02631-w

Maranna, S., Verma, K., Talukdar, A., Lal, S. K., Kumar, A., & Mukherjee, K. (2016). Introgression of null allele of Kunitz trypsin inhibitor through marker-assisted backcross breeding in soybean (*Glycine max* L. Merr.). *BMC Genetics*, *17*, 106. https://doi.org/10.1186/s12863-016-0413-2

Marsh, J. I., Hu, H., Petereit, J., Bayer, P. E., Valliyodan, B., Batley, J., Nguyen, H. T., & Edwards, D. (2022). Haplotype mapping uncovers unexplored variation in wild and domesticated soybean at the major protein locus cqProt-003. *Theoretical and Applied Genetics*, *135*, 1443–1455. https://doi.org/10.1007/s00122-022-04045-8

Martino, H. S. D., Martin, B. R., Weaver, C. M., Bressan, J., Moreira, M. A., & Costa, N. M. B. (2008). A soybean cultivar lacking lipoxygenase 2 and 3 has similar calcium bioavailability to a commercial variety despite higher calcium absorption inhibitors. *Journal of Food Science*, *73*, H33–H35. https://doi.org/10.1111/j.1750-3841.2008.00684.x

Massman, J. M., Jung, H.-J. G., & Bernardo, R. (2013). Genome-wide selection versus marker-assisted recurrent selection to improve grain yield and stover-quality traits for cellulosic ethanol in maize. *Crop Science*, *53*, 58–66. https://doi.org/10.2135/cropsci2012.02.0112

Mattick, L. R., & Hand, D. B. (1969). Identification of a volatile component in soybeans that contributes to the raw bean flavor. *Journal of Agricultural and Food Chemistry*, *17*, 15–17. https://doi.org/10.1021/jf60161a006

Maughan, P. J., Saghai Maroof, M. A., & Buss, G. R. (2000). Identification of quantitative trait loci controlling sucrose content in soybean (*Glycine max*). *Molecular Breeding*, *6*, 105–111. https://doi.org/10.1023/A:1009628

McCarty, N. (2020). *CRISPR—How it works, top applications and how to use it yourself*. Medium. https://nikomccarty.medium.com/almost-everything-you-should-know-about-crispr-how-it-works-top-applications-and-how-to-use-it-61e27b04bea6

Medic, J., Atkinson, C., & Hurburgh, C. R., Jr. (2014). Current knowledge in soybean composition. *Journal of the American Oil Chemists' Society*, *91*, 363–384. https://doi.org/10.1007/s11746-013-2407-9

Meksem, K., Njiti, V. N., Banz, W. J., Iqbal, M. J., Kassem, M. M., Hyten, D. L., Yuang, J., Winters, T. A., & Lightfoot, D. A. (2001). Genomic regions that underlie soybean seed isoflavone content. *BioMed Research International*, *1*, 38–44. https://doi.org/10.1155/S1110724301000110

Meng, S., Chang, S., Gillen, A. M., & Zhang, Y. (2016). Protein and quality analyses of accessions from the USDA soybean germplasm collection for tofu production. *Food Chemistry*, *213*, 31–39. https://doi.org/10.1016/j.foodchem.2016.06.046

Messina, M. J. (1999). Legumes and soybeans: Overview of their nutritional profiles and health effects. *The American Journal of Clinical Nutrition*, *70*(3), 439s–450s. https://doi.org/10.1093/ajcn/70.3.439s

Messina, M., & Messina, V. (2010). The role of soy in vegetarian diets. *Nutrients*, *2*, 855–888. https://doi.org/10.3390/nu2080855

Meuwissen, T. H. E., Hayes, B. J., & Goddard, M. E. (2001). Prediction of total genetic value using genome-wide dense marker maps. *Genetics*, *157*(4), 1819–1829. https://doi.org/10.1093/genetics/157.4.1819

Miladinović, J., Đorđević, V., Balešević-Tubić, S., Petrović, K., Ćeran, M., Cvejić, J., Bursać, M., & Miladinović, D. (2019). Increase of isoflavones in the aglycone form in soybeans by targeted crossings of cultivated breeding material. *Scientific Reports*, *9*, 10341. https://doi.org/10.1038/s41598-019-46817-1

Moser, B. R., Haas, M. J., Winkler, J. K., Jackson, M. A., Erhan, S. Z., & List, G. R. (2007). Evaluation of partially hydrogenated methyl esters of soybean oil as biodiesel. *European Journal of Lipid Science and Technology*, *109*(1), 17–24. https://doi.org/10.1002/ejlt.200600215

Mourtzinis, S., Gaspar, A. P., Naeve, S. L., & Conley, S. P. (2017). Planting date, maturity, and temperature effects on soybean seed yield and composition. *Crop Ecology and Physiology*, *109*(5), 2040–2049. https://doi.org/10.2134/agronj2017.05.0247

Mullin, W. J., & Xu, W. (2001). Study of soybean seed coat components and their relationship to water absorption. *Journal of Agricultural and Food Chemistry*, *49*, 5331–5335. https://doi.org/10.1021/jf010303s

Murphy, P. A., Chen, H. P., Hauck, C. C., & Wilson, L. A. (1997). Soybean protein composition and tofu quality. *Food Technology*, *51*(3), 86–88.

Nakasathien, S. I., Israel, D. W., Wilson, R. F., & Kwanyuen, P. (2000). Regulation of seed protein concentration by supra-optimal nitrogen supply. *Crop Science*, *40*, 1277–1284. https://doi.org/10.2135/cropsci2000.4051277x

Nekrasov, V., Wang, C., Win, J., Lanz, C., Weigel, D., & Kamoun, S. (2017). Rapid generation of a transgene-free powdery mildew resistant tomato by genome deletion. *Scientific Reports*, *7*, 482. https://doi.org/10.1038/s41598-017-00578-x

Nielsen, N., Dickinson, C., Cho, T.-J. H., Thanh, V. J., Scallon, B. L., Fischer, R., Sims, T. N., Drews, G. B., & Goldberg, R. (1989). Characterization of glycinin gene family in soybean. *Plant Cell*, *1*, 313–328. https://doi.org/10.1105/tpc.1.3.313

Nill, K. (2016). *Soy beans: Properties and analysis*. US Soybean Export Council. http://doi.org/10.1016/B978-0-12-384947-2.00642-5

Obendorf, R. L., Zimmerman, A. D., Zhang, Q., Castillo, A., Kosina, S. M., Bryant, E. G., Sensening, E. M., Wu, J., & Schnebly, S. R. (2009). Accumulation of soluble carbohydrates during seed development and maturation of low-raffinose, low-stachyose soybean. *Crop Science*, *49*(1), 329–341. https://doi.org/10.2135/cropsci2008.06.0370

Obendorf, R. L. (1997). Oligosaccharides and galactosyl cyclitols in seed desiccation tolerance. *Seed Science Research*, *7*, 63–74. https://doi.org/10.1017/S096025850000341X

Ogawa, T., Samato, M., & Takahashi, K. (2000). Soybean allergens and hypoallergenic soybean products. *Journal of Nutritional Science and Vitaminology*, *46*(6), 271–279. https://doi.org/10.3177/jnsv.46.271

Okeke, U. G., Akdemir, D., Rabbi, I., Kulakow, P., & Jannink, J.-L. (2017). Accuracies of univariate and multivariate genomic prediction models in African cassava. *Genetics Selection Evolution*, *49*, 88. https://doi.org/10.1186/s12711-017-0361-y

Orazaly, M., Chen, P., Zhang, B., Florez-Palacios, L., & Zeng, A. (2018). Confirmation of SSR Markers and QTL for seed calcium content and hardness of soybean. *Journal of Crop Improvement*, *32*, 71–89. https://doi.org/10.1080/15427528.2017.1398701

Orf, J. H., Chase, K., Jarvik, T., Mansur, L. M., Cregan, P. B., Adler, F. R., & Lark, K. G. (1999). Genetics of soybean agronomic traits: I. comparison of three related recombinant inbred populations. *Crop Science*, *39*(6), 1642–1651. https://doi.org/10.2135/cropsci1999.3961642x

Ostezan, A., Prenger, E. M., Rosso, L., Zhang, B., Stupar, R. M., Glenn, T., Mian, M. A. R., & Li, Z. (2023). A chromosome 16 deletion conferring a high sucrose phenotype in soybean. *Theoretical and Applied Genetics*, *136*, 109. https://doi.org/10.1007/s00122-023-04354-6

Paine, J. A., Shipton, C. A., Chaggar, S., Howells, R. M., Kennedy, M. J., Vernon, G., Wright, S. Y., Hinchliffe, E., Adams, J. L., Silverstone, A. L., & Drake, R. (2005). Improving the nutritional value of Golden Rice through increased pro-vitamin A content. *Nature Biotechnology*, *23*, 482–487. https://doi.org/10.1038/nbt1082

Pantalone, V., & Smallwood, C. (2018). Registration of ‘TN11-5102’ soybean cultivar with high yield and high protein meal. *Journal of Plant Registrations*, *12*, 304–308. https://doi.org/10.3198/jpr2017.10.0074crc

Pantalone, V., Smallwood, C., & Fallen, B. (2017). Development of ‘Ellis’ soybean with high soymeal protein, resistance to stem canker, southern root knot nematode, and frogeye leaf spot. *Journal of Plant Registrations*, *11*, 250–255. https://doi.org/10.3198/jpr2016.12.0071crc

Pantalone, V. R., Wilson, R. F., Novitzky, W. P., & Burton, J. W. (2002). Genetic regulation of elevated stearic acid concentration in soybean oil. *Journal of the American Oil Chemists Society*, *79*(6), 549–553. https://doi.org/10.1007/s11746-002-0520-8

Panthee, D. R., Kwanyuen, P., Sams, C. E., West, D. R., Saxton, A. M., & Pantalone, V. R. (2004). Quantitative trait loci for β-conglycinin (7S) and glycinin (11S) fractions of soybean storage protein. *Journal of the American Oil Chemists’ Society*, *81*, 1005–1012. https://doi.org/10.1007/s11746-004-1014-4

Panthee, D. R., Pantalone, V. R., Sams, C. E., Saxton, A. M., West, D. R., Orf, J. H., & Killam, A. S. (2006). Quantitative trait loci controlling sulfur containing amino acids, methionine and cysteine, in soybean seeds. *Theoretical and Applied Genetics*, *112*(3), 546–553. https://doi.org/10.1007/s00122-005-0161-6

Panthee, D. R., Pantalone, V. R., & Saxton, A. M. (2006). Modifier QTL for fatty acid composition in soybean oil. *Euphytica*, *152*(1), 67–73. https://doi.org/10.1007/s10681-006-9179-3

Panthee, D. R., Pantalone, V. R., West, D. R., Saxton, A. M., & Sams, C. E. (2005). Quantitative trait loci for seed protein and oil concentration, and seed size in soybean. *Crop Science*, *45*(5), 2015–2022. https://doi.org/10.2135/cropsci2004.0720

Parsons, C. M., Zhang, Y., & Araba, M. (2000). Nutritional evaluation of soybean meals varying in oligosaccharide content. *Poultry Science*, *79*(8), 1127–1131. https://doi.org/10.1093/ps/79.8.1127

Patil, G., Mian, R., Vuong, T., Pantalone, V., Song, Q., Chen, P., Shannon, G. J., Carter, T. C., & Nguyen, H. T. (2017). Molecular mapping and genomics of soybean seed protein, a review and perspective for the future. *Theoretical and Applied Genetics*, *130*, 1975–1991. https://doi.org/10.1007/s00122-017-2955-8

Paz, M. M., Shou, H., Guo, Z., Zhang, Z., Banerjee, A. K., & Wang, K. (2004). Assessment of conditions affecting Agrobacterium-mediated soybean transformation using the cotyledonary node explant. *Euphytica*, *136*(2), 167–179. https://doi.org/10.1023/B:EUPH.0000030669.75809.dc

Pazdernik, D. L., Killam, A. S., & Orf, J. H. (1997). Analysis of amino and fatty acid composition in soybean seed, using near infrared reflectance spectroscopy. *Agronomy Journal*, *89*(4), 679–685. https://doi.org/10.2134/agronj1997.00021962008900040022x

Pereira, L. A. (2020). *Modifying soluble carbohydrates in soybean seed for enhanced nutritional energy meal*. Soybean Research & Information Network. https://soybeanresearchinfo.com/research-highlight/modifying-soluble-carbohydrates-in-soybean-seed-for-enhanced-nutritional-energy-meal/

Peterbauer, T. P., & Richter, A. (2001). Biochemistry and physiology of raffinose family oligosaccharides and galactosyl cyclitols in seeds. *Seed Science Research*, *11*, 185–197. https://doi.org/10.1079/SSR200175

Pettigrew, J. E., Soltwedel, K. T., Miguel, J. C., & Palacios, M. F. (2002). *Soybean use- swine. soybean meal information center fact sheet*. The Soybean Meal Information Center, United Soybean Board.

Pfarr, M. D., Kazula, M. J., Miller-Garvin, J. E., & Naeve, S. K. (2018). Amino acid balance is affected by protein concentration in soybean. *Crop Science*, *58*, 2050–2062. https://doi.org/10.2135/cropsci2017.11.0703

Pham, A.-T., Lee, J.-D., Shannon, J. G., & Bilyeu, K. D. (2010). Mutant alleles of FAD2-1A and FAD2-1B combine to produce soybeans with the high oleic acid seed oil trait. *BMC Plant Biology*, *10*, 195. https://doi.org/10.1186/1471-2229-10-195

Pham, A.-T., Shannon, J. G., & Bilyeu, K. D. (2012). Combinations of mutant FAD2 and FAD3 genes to produce high oleic acid and low linolenic acid soybean oil. *Theoretical and Applied Genetics*, *125*(3), 503–515. https://doi.org/10.1007/s00122-012-1849-z

Phillips, R. D. (1993). Starchy legumes in human nutrition, health and culture. *Plant Foods for Human Nutrition*, *44*, 195–211. https://doi.org/10.1007/BF01088314

Piper, E. L., & Boote, K. I. (1999). Temperature and cultivar effects on soybean seed oil and protein concentrations. *Journal of American Oil Chemists' Society*, *76*, 1233–1241. https://doi.org/10.1007/s11746-999-0099-y

Pipolo, A., Sinclair, T. R., & Camara, G. M. S. (2004). Effects of temperature on oil and protein concentration in soybean seeds cultured *in vitro*. *Annals of Applied Biology*, *144*(1), 71–76. https://doi.org/10.1111/j.1744-7348.2004.tb00318.x

Qi, Z., Wu, Q., Han, X., Sun, Y., Du, X., Liu, C., Jiang, H., Hu, G., & Chen, Q. (2011). Soybean oil content QTL mapping and integrating with meta-analysis method for mining genes. *Euphytica*, *179*(3), 499–514. https://doi.org/10.1007/s10681-011-0386-1

Qin, J., Shi, A., Song, Q., Li, S., Wang, F., Cao, Y., Ravelombola, W., Song, Q., Yang, Y., & Zhang, M. (2019). Genome wide association study and genomic selection of amino acid concentrations in soybean seeds. *Frontiers in Plant Science*, *10*, 1445. https://doi.org/10.3389/fpls.2019.01445

Qin, P., Wang, T., & Luo, Y. (2022). A review on plant-based proteins from soybean: Health benefits and soy product development. *Journal of Agriculture and Food Research*, *7*, 100265. https://doi.org/10.1016/j.jafr.2021.100265

QSorter Technology. (n.d.). QualySense. https://www.qualysense.com/howitworks

Raboy, V. (2002). Progress in breeding low phytate crops. *The Journal of Nutrition*, *132*(3), 503S–5055S. https://doi.org/10.1093/jn/132.3.503S

Radcliffe, J. S., Brito, L. F., Reddivari, L., Schmidt, M., Herman, E. M., & Schinckel, A. P. (2019). A swine model of soy protein–induced food allergenicity: Implications in human and swine nutrition. *Animal Frontiers*, *9*, 52–59. https://doi.org/10.1093/af/vfz025

Ravelombola, W. S., Qin, J., Shi, A., Nice, L., Bao, Y., Lorenz, A., Orf, J. H., Young, N. D., & Chen, S. (2019). Genome-wide association study and genomic selection for soybean chlorophyll content associated with soybean cyst nematode tolerance. *BMC Genomics*, *20*, 904. https://doi.org/10.1186/s12864-019-6275-z

Ravelombola, W., Qin, J., Shi, A., Song, Q., Yuan, J., Wang, F., Chen, P., Yan, L., Feng, Y., Zhao, T., Meng, Y., Guan, K., Yang, C., & Zhang, M. (2021). Genome-wide association study and genomic selection for yield and related traits in soybean. *PLoS One*, *16*(8), e0255761. https://doi.org/10.1371/journal.pone.0255761

Ray, J. D., Fritschi, F. B., & Heatherly, L. G. (2006). Large applications of fertilizer N at planting affects seed protein and oil concentration and yield in the early soybean production system. *Field Crops Research*, *99*, 67–74. https://doi.org/10.1016/j.fcr.2006.03.006

Reinprecht, Y., Poysa, V. W., Yu, K., Rajcan, I., Ablett, G. R., & Pauls, K. P. (2006). Seed and agronomic QTL in low linolenic acid, lipoxygenase-free soybean (G*lycine max* (L.) Merr.) germplasm. *Genome*, *49*(12), 1510–1527. https://doi.org/10.1139/g06-112

Ren, C., Bilyeu, K. D., & Beuselinck, P. (2009). Composition, vigor, and proteome of mature soybean seeds developed under high temperature. *Crop Science*, *49*, 1010–1022. https://doi.org/10.2135/cropsci2008.05.0247

Rennie, B. D., Zilka, J., Cramer, M. M., & Beversdorf, W. D. (1988). Genetic analysis of low linolenic acid levels in the soybean line PI 361088B. *Crop Science*, *28*(4), 655–657. https://doi.org/10.2135/cropsci1988.0011183X002800040019x

Riaz, A., Raza, Q., Kumar, A., Dean, D., Chiwina, K., Phiri, T. M., Thomas, J., & Shi, A. (2023). GWAS and genomic selection for marker-assisted development of sucrose enriched soybean cultivars. *Euphytica*, *219*, 97. https://doi.org/10.1007/s10681-023-03224-y

Rincker, K., Nelson, R., Specht, J., Sleper, D., Cary, T., Cianzio, S. R., Casteel, S., Conley, S., Chen, P., Davis, V., Fox, C., Graef, G., Godsey, C., Holshouser, D., Jiang, G.-L., Kantartzi, S. K., Kenworthy, W., Lee, C., Mian, R., … Diers, B. (2014). Genetics improvement of U.S. soybean in maturity group II, III, and IV. *Crop Science*, *54*(4), 1419–1432. https://doi.org/10.2135/cropsci2013.10.0665

Rogers, J., Chen, P., Shi, A., Zhang, B., Scaboo, A., Smith, S. F., & Zeng, A. (2015). Agronomic performance and genetic progress of selected historical soybean varieties in the southern USA. *Plant Breeding*, *134*, 85–93. https://doi.org/10.1111/pbr.12222

Rosset, M., Prudencio, S. H., & Beleia, A. D. P. (2012). Viscozyme L action on soy slurry affects carbohydrates and antioxidant properties of silken tofu. *Food Science and Technology International*, *18*, 531–538. https://doi.org/10.1177/1082013211433076

Rosso, M. L., Burleson, S. A., Maupin, L. M., & Rainey, K. M. (2011). Development of breeder-friendly markers for selection of MIPS1 mutations in soybean. *Molecular Breeding*, *28*, 127–132. https://doi.org/10.1007/s11032-011-9573-y

Rosso, M. L., Shang, C., Song, Q., Escamilla, D., Gillenwater, J., & Zhang, B. (2021). Development of breeder-friendly KASP markers for low concentration of Kunitz trypsin inhibitor in soybean seeds. *International Journal of Molecular Sciences*, *22*, 2675. https://doi.org/10.3390/ijms22052675

Rotundo, J. L., & Westgate, M. E. (2009). Meta-analysis of environmental effects on soybean seed composition. *Field Crops Research*, *110*, 147–156. https://doi.org/10.1016/j.fcr.2008.07.012

Rotundo, J. L., & Westgate, M. E. (2010). Rate and duration of seed component accumulation in water-stressed soybean. *Crop Science*, *50*(2), 676–684. https://doi.org/10.2135/cropsci2009.05.0240

Saghai Maroof, M. A., & Buss, G. R. (2008). *Low phytic acid, low stachyose, high sucrose soybean lines* (U.S. Patent No. 12/033,830). U.S. Patent and Trademark Office.

Saio, K. (1976). Soybeans resistant to water absorption. *Cereal Foods World*, *21*, 168–173.

Saio, K. (1979). Tofu-relationships between texture and fine structures. *Cereal Foods World*, *24*(8), 342–354.

Saio, K., Kamiya, M., & Watanabe, T. (1969). Food processing characteristics of soybean 11S and 7S proteins. *Agricultural and Biological Chemistry*, *33*(9), 1303–1308. https://doi.org/10.1080/00021369.1969.10859465

Saio, K., & Watanabe, T. (1978). Differences in functional properties of 7S and 11S soybean protein. *Journal of Texture Studies*, *9*(1–2), 135–157. https://doi.org/10.1111/j.1745-4603.1978.tb01298.x

Sato, T., Equchi, K., Hatano, T., & Nishiba, Y. (2008). Use of Near-infrared Reflectance Spectroscopy for the estimation of the isoflavone contents of soybean seeds. *Plant Production Science*, *11*(4), 481–486. https://doi.org/10.1626/pps.11.481

Scaboo, A. M., Pantalone, V. R., Walker, D. R., Boerma, H. R., West, D. R., Walker, F. R., & Sams, C. R. (2009). Confirmation of molecular markers and agronomic traits associated with seed phytate content in two soybean RIL populations. *Crop Science*, *49*, 426–432. https://doi.org/10.2135/cropsci2007.11.0614

Scallon, B., Thanh, V., Floener, L., & Nielsen, N. (1985). Identification and characterization of DNA clones encoding group-II glycinin subunits. *Theoretical and Applied Genetics*, *70*, 510–519. https://doi.org/10.1007/BF00305984

Schaeffer, L. R. (2006). Strategy for applying genome-wide selection in dairy cattle. *Journal of Animal Breeding and Genetics*, *123*(4), 218–223. https://doi.org/10.1111/j.1439-0388.2006.00595.x

Schlueter, J. A., Vasylenko-Sanders, I. F., Deshpande, S., Yi, J., Siegfried, M., roe, B. A., Schlueter, S. D., Scheffler, B. E., & Shoemaker, R. C. (2007). The FAD2 gene family of soybean: Insights into the structural and functional divergence of a paleopolyploid genome. *Crop Science*, *47*, S14–S-26. https://doi.org/10.2135/cropsci2006.06.0382tpg

Scott, W. O., & Aldrich, S. R. (1970). *Modern soybean production* (1st ed.) S & A Publications.

Sebastian, S. A., Kerr, P. S., Pearlstein, R. W., & Hitz, W. D. (2000). Soybean germplasm with novel genes for improved digestibility. In J. K. Drackley (Ed.), *Soy in animal nutrition* (pp. 56–73). Federation of Animal Science Societies.

Shannon, J. G., Novitzky, W. P., Sleper, D. A., Burton, J. W., Chappell, A. S., & Anand, S. C. (2007). Registration of S01-9364 and S01-9391 soybean germplasm resistant to soybean cyst nematode with seed oil low in linolenic acid. *Journal of Plant Registrations*, *1*, 164–165. https://doi.org/10.3198/jpr2006.07.0454crg

Shen, B., Schmidt, M. A., Collet, K. H., Liu, Z.-B., Coy, M., Abbitt, S., Molloy, L., Frank, M., Everard, J. D., Booth, R., Samadar, P. P., He, Y., Kinney, A., & Herman, E. M. (2022). RNAi and CRISPR–Cas silencing E3-RING ubiquitin ligase AIP2 enhances soybean seed protein content. *Journal of Experimental Botany*, *73*(22), 7285–7297. https://doi.org/10.1093/jxb/erac376.org/10.1111/j.1753-4887.2008.00106.x

Shi, J., Gao, H., Wang, H., Lafitte, H. R., Archibald, R. L., Yang, M., Hakimi, S. M., Mo, H., & Habben, J. E. (2017). ARGOS8 variants generated by CRISPR-Cas9 improve maize grain yield under field drought stress conditions. *Plant Biotechnology Journal*, *15*(2), 207–216. https://doi.org/10.1111/pbi.12603

Shi, M.-Q., Liao, X.-L., Ye, Q., Zhang, W., Li, Y.-K., Bhat, J. A., Kan, G.-Z., & Yu, D.-Y. (2022). Linkage and association mapping of wild soybean (Glycine soja) seeds germinating under salt stress. *Journal of Integrative Agriculture*, *21*(10), 2833–2847. https://doi.org/10.1016/j.jia.2022.07.031

Shorter, R., Byth, D. E., & Mungomery, V. E. (1977). Estimates of selection parameters associated with protein and oil content of soybean seeds (*Glycine max* (L.) Merr.). *Australian Journal of Agricultural Research*, *28*, 211–222.

Singer, W. M., Shea, Z., Yu, D., Huang, H., Mian, M. A. R., Shang, C., Rosso, M. L., Song, Q. J., & Zhang, B. (2022). Genome-wide association study and genomic selection for proteinogenic methionine in soybean seeds. *Frontiers in Plant Science*, *13*, 859109. https://doi.org/10.3389/fpls.2022.859109

Singh, A., Meena, M., Kumar, D., Dubey, A. K., & Hassan, M. I. (2015). Structural and functional analysis of various globulin proteins from soy seed. *Critical Reviews in Food Science and Nutrition*, *55*(11), 1491–1502. https://doi.org/10.1080/10408398.2012.700340

Skoneczka, J. A., Maroof, M. S., Shang, C., & Buss, G. R. (2009). Identification of candidate gene mutation associated with low stachyose phenotype in soybean line PI200508.*Crop Science*, *49*(1), 247–255. https://doi.org/10.2135/cropsci2008.07.0403

Smallwood, C. J., Saxton, A. M., Gillman, J. D., Bhandari, H. S., Wadl, P. A., Fallen, B. D., Hyten, D. L., Song, Q., & Pantalone, V. R. (2019). Context-specific genomic selection strategies outperform phenotypic selection for soybean quantitative traits in the progeny row stage. *Crop Science*, *59*(1), 54–67. https://doi.org/10.2135/cropsci2018.03.0197

Sonah, H., O'Donoughue, L., Cober, E., Rajcan, I., & Belzile, F. (2015). Identification of loci governing eight agronomic traits using a GBS-GWAS approach and validation by QTL mapping in soya bean. *Plant Biotechnology Journal*, *13*(2), 211–221. https://doi.org/10.1111/pbi.12249

Song, Q., Hyten, D. L., Jia, G., Quigley, C. V., Fickus, E. W., Nelson, R. L., & Cregan, P. B. (2013). Development and evaluation of SoySNP50K, a high-density genotyping array for soybean. *PLoS ONE*, *8*(1), e54985. https://doi.org/10.1371/journal.pone.0054985

Song, Q., Yan, L., Quigley, C., Fickus, E., Wei, H., Chen, L., Dong, F., Araya, S., Liu, J., Hyten, D., Pantalone, V., & Nelson, R. L. (2020). Soybean BARCSoySNP6K: An assay for soybean genetics and breeding research. *Plant Journal*, *104*(3), 800–811. https://doi.org/10.1111/tpj.14960

Soy Stats. (2022). *A reference guide to important soybean facts and figures*. The American Soybean Association. http://soystats.com/wp-content/uploads/Soy-Stats-2022-web.pdf

Specht, J., & Williams, J. H. (1984). Contribution of genetic technology to soybean productivity—Retrospect and prospect. In W. R. Fehr (Ed.), *Genetic contributions to yield gains of five major crop plants* (pp. 49–74). Crop Science Society of America, American Society of Agronomy.

Stewart-Brown, B. B., Song, Q., Vaughn, J. N., & Li, Z. (2019). Genomic selection for yield and seed composition traits within an applied soybean breeding program. *G3: Genes, Genomes, Genetics*, *9*(7), 2253–2265. https://doi.org/10.1534/g3.118.200917

Sudarić, A., ŠImić, D., & Vratarić, M. (2006). Characterization of genotype by environment interactions in soybean breeding programmes of southeast Europe. *Plant Breeding*, *125*(2), 191–194. https://doi.org/10.1111/j.1439-0523.2006.01185.x

Sugano, S., Hirose, A., Kanazashi, Y., Adachi, K., Hibara, M., Itoh, T., Mikami, M., Endo, M., Hirose, S., Maruyama, N., Abe, J., & Yamada, T. (2020). Simultaneous induction of mutant alleles of two allergenic genes in soybean by using site-directed mutagenesis. *BMC Plant Biology*, *20*, 513. https://doi.org/10.1186/s12870-020-02708-6

Sui, M., Wang, Y., Bao, Y., Wang, X., Li, R., Lv, Y., Yan, M., Quan, C., Li, C., Teng, W., Li, W., Zhao, X., & Han, Y. (2020). Genome-wide association analysis of sucrose concentration in soybean (*Glycine max* L.) seed based on high-throughput sequencing. *Plant Genome*, *13*(3), e20059. https://doi.org/10.1002/tpg2.20059

Synthego. (n.d.). *CRISPR in agriculture: An era of food evolution*. https://www.synthego.com/blog/crispr-agriculture-foods

Tang, G.-Q., Novitzky, W. P., Carol Griffin, H., Huber, S. C., & Dewey, R. E. (2005). Oleate desaturase enzymes of soybean: Evidence of regulation through differential stability and phosphorylation. *The Plant Journal*, *44*(3), 433–446. https://doi.org/10.1111/j.1365-313X.2005.02535.x

Teraishi, M., Takahashi, M., Hajika, M., Matsunaga, R., Uematsu, Y., & Ishimoto, M. (2001). Suppression of soybean β-conglycinin genes by a dominant gene, *Scg-1*. *Theoretical and Applied Genetics*, *103*, 1266–1272. https://doi.org/10.1007/s001220100702

Thakur, M., & Hurburgh, C. R. (2007). Quality of US soybean meal compared to the quality of soybean meal from other origins. *Journal of* the American Oil Chemists' Society*, *84*, 835–843. https://doi.org/10.1007/s11746-007-1107-8

Thanh, V. H., & Shibasaki, K. (1978). Major proteins of soybean seeds. subunit structure of β- conglycinin. *Journal of Agricultural and Food Chemistry*, *26*, 692–695. https://doi.org/10.1021/jf60217a026

Thompson, M. M., Niemuth, A., Sabbatini, J., Levin, D., Breeze, M. L., Li, X., Perez, T., Taylor, M., & Harrigan, G. G. (2016). Analysis of vitamin K1 in soybean seed: Assessing levels in a lineage representing over 35 years of breeding. *Journal of the American Oil Chemists' Society*, *93*(4), 587–594. https://doi.org/10.1007/s11746-016-2795-8

Tsubokura, Y., Hajika, M., & Harada, K. (2006). Molecular markers associated with β-conglycinin deficiency in soybean. *Breeding Science*, *56*, 113–117. https://doi.org/10.1270/jsbbs.56.113

United Soybean Board (USB). (2018). *Meal protein and amino acid considerations: Crop year 2018*. https://api.unitedsoybean.org/uploads/documents/6-usb-ussm-2018-meal-protein-and-amino-acid-considerations.pdf

United Soybean Board (USB). (2022). *Value of U.S. soybeans: The proof is in the nutrients*. https://www.unitedsoybean.org/hopper/value-of-u-s-soybeans-the-proof-is-in-the-nutrients/

United States Department of Agriculture (USDA). (2013). *Oilseeds: World markets and trade*. https://downloads.usda.library.cornell.edu/usda-esmis/files/tx31qh68h/2514nk855/6395w7441/oilseed-trade-12-10-2013.pdf

United States Department of Agriculture (USDA). (2023). *Oilseeds: World markets and trade*. https://downloads.usda.library.cornell.edu/usda-esmis/files/tx31qh68h/02872f03s/8910mb76h/oilseeds.pdf

United States Soybean Export Council (USSEC). (2022). *Conversion table*. https://ussec.org/resources/conversion-table/

Utsumi, S., & Kinsella, J. E. (1985). Forces involved in soy protein gelation: Effects of various reagents on the formation, hardness, and solubility of heat-induced gels made from 7S, 11S, and soy isolate. *Journal of Food Science*, *50*, 1278–1282. https://doi.org/10.1111/j.1365-2621.1985.tb10461.x

Valentine, M. F., De Tar, J. R., Mookkan, M., Firman, J. D., & Zhang, Z. J. (2017). Silencing of soybean raffinose synthase gene reduced raffinose family oligosaccharides and increased true metabolizable energy of poultry feed. *Frontiers in Plant Science*, *8*, 692. https://doi.org/10.3389/fpls.2017.00692

van den Heuvel, A. (2021). *ME-PRO® platinum: The effect and presence of anti-nutritional factors in soy products*. E.F.S.-Holland. https://www.efs-holland.com/news/me-pro-the-effect-and-presence-of-anti-nutritional-factors-in-soy-products/

van der Ven, C., Matser, A. M., & van den Berg, R. W. (2005). Inactivation of soybean trypsin inhibitors and lipoxygenase by high-pressure processing. *Journal of Agricultural and Food Chemistry*, *53*, 1087–1092. https://doi.org/10.1021/jf048577d

Vaughn, J. N., Nelson, R. L., Song, Q., Cregan, P. B., & Li, Z. (2014). The genetic architecture of seed composition in soybean is refined by genome-wide association scans across multiple populations. *G3: Genes, Genomes, Genetics*, *4*(11), 2283–2294. https://doi.org/10.1534/g3.114.013433

Vieira, C. C., & Chen, P. (2021). The numbers game of soybean breeding in the United States. *Crop Breeding and Applied Biotechnology*, *21*, 387521–387531. https://doi.org/10.1590/1984

Vieira, C. C., Persa, R., Chen, P., & Jarquin, D. (2022). Incorporation of soil-derived covariates in progeny testing and line selection to enhance genomic prediction accuracy in soybean breeding. *Frontiers in Genetics*, *13*, 905824. https://doi.org/10.3389/fgene.2022.905824

Voldeng, H. D., Cober, E. R., Hume, D. J., Gillard, C., & Morrison, M. J. (1997). Fifty-eight years of genetic improvement of short-season soybean cultivars in Canada. *Crop Science*, *37*(2), 428–431. https://doi.org/10.2135/cropsci1997.0011183X003700020020x

Vollmann, J., Fritz, C. N., Wagentristl, H., & Ruckenbauer, P. (2000). Environmental and genetic variation of soybean seed protein content under central European growing conditions. *Journal of the Science of Food and Agriculture*, *80*, 1300–1306. https://doi.org/10.1002/1097-0010(200007)80:9%3C1300::AID-JSFA640%3E3.0.CO;2-I

Walker, D. R., Scaboo, A. M., Pantalone, V. R., Wilcox, J. R., & Boerma, H. R. (2006). Genetic mapping of loci associated with seed phytic acid content in CX1834-1-2 soybean. *Crop Science*, *46*, 390–397. https://doi.org/10.2135/cropsci2005.0245

Wang, B., Zhang, Q., Zhang, N., Bak, K. H., Soladoye, O. P., Aluko, R. E., Fu, Y., & Zhang, Y. (2021). Insights into formation, detection and removal of the beany flavor in soybean protein. *Trends in Food Science & Technology*, *112*, 336–347. https://doi.org/10.1016/j.tifs.2021.04.018

Wang, J., Chen, P., Wang, D., Shannon, G., Zeng, A., Orazaly, M., & Wu, C. (2015). Identification and mapping of stable QTL for protein content in soybean seeds. *Molecular Breeding*, *35*, 92. https://doi.org/10.1007/s11032-015-0285-6

Wang, J., Kuang, H., Zhang, Z., Yang, Y., Yan, L., Zhang, M., Song, S., & Guan, Y. (2020). Generation of seed lipoxygenase-free soybean using CRISPR-Cas9. *Crop Journal*, *8*, 432–439. https://doi.org/10.1016/j.cj.2019.08.008

Wang, J., Liu, L., Guo, Y., Wang, Y.-H., Zhang, L., Jin, L.-G., Guan, R.-X., Liu, Z.-X., Wang, L.-L., Chang, R.-Z., & Qiu, L.-J. (2014). A dominant locus, qBSC-1, controls β subunit content of seed storage protein in soybean (*Glycine max* (L.) Merri.). *Journal of Integrative Agriculture*, *13*(9), 1854–1864. https://doi.org/10.1016/S2095-3119(13)60579-1

Wang, Q., Ge, X., Tian, X., Zhang, Y., Zhang, J., & Zhang, P. (2013). Soy isoflavone: The multipurpose phytochemical (Review). *Biomedical Reports*, *1*(5), 697–701. https://doi.org/10.3892/br.2013.1291

Wang, Y., Chen, P., & Zhang, B. (2014). Quantitative trait loci analysis of soluble sugar contents in soybean. *Plant Breeding*, *133*(4), 493–498. https://doi.org/10.1111/pbr.12178

Wang, Z., Shea, Z., Rosso, L., Shang, C., Li, J., Bewick, P., Li, Q., Zhao, B., & Zhang, B. (2022). Development of molecular markers of the Kunitz trypsin inhibitor mutant alleles generated by CRISPR/Cas9-mediated mutagenesis in soybean. bioRxiv. https://doi.org/10.1101/2022.08.22.504807

Wang, Z., Yang, M., Wang, Y., Yang, C., Zhao, X., Tian, L., Shao, Y., Bao, G., Huang, J., Xia, N., & Han, Y. (2020). Genetic dissection of 2-heptenal content in soybean (*Glycine max*) seed through genome-wide association study. *Crop and Pasture Science*, *71*, 884–893. https://doi.org/10.1071/CP20237

Wang, Z., Yu, D., Morota, G., Dhakal, K., Singer, W., Lord, N., Huang, H., Chen, P., Mozzoni, L., Li, S., & Zhang, B. (2023). Genome-wide association analysis of sucrose and alanine contents in edamame beans. *Frontiers in Plant Science*, *13*, 1086007. https://doi.org/10.3389/fpls.2022.1086007

Warrington, C. V., Abdel-Haleem, H., Hyten, D. L., Cregan, P. B., Orf, J. H., Killam, A. S., Bajjalieh, N., Li, Z., & Boerma, H. R. (2015). QTL for seed protein and amino acids in the Benning × Danbaekkong soybean population. *Theoretical and Applied Genetics*, *128*(5), 839–850. https://doi.org/10.1007/s00122-015-2474-4

Warrington, C. V., Abdel-Haleem, H., Orf, J. H., Killam, A. S., Bajjaleih, N., Li, Z., & Boerma, H. R. (2014). Resource allocation for selection of seed protein and amino acids in soybean. *Crop Science*, *54*(3), 963–970. https://doi.org/10.2135/cropsci2013.12.0799

Wartha, C. A., & Lorenz, A. J. (2021). Implementation of genomic selection in public-sector plant breeding programs: Current status and opportunities. *Crop Breeding and Applied Biotechnology*, *21*, e394621S15. https://doi.org/10.1590/1984-70332021v21Sa28

Watanabe, D., Adányi, N., Takács, K., Maczó, A., Nagy, A., Gelencsér, É., Pachner, M., Lauter, K., Baumgartner, S., & Vollmann, J. (2017). Development of soybeans with low P34 allergen protein concentration for reduced allergenicity of soy foods. *Journal of the Science of Food and Agriculture*, *97*, 1010–1017. https://doi.org/10.1002/jsfa.7827

Weber, C. R., & Fehr, W. R. (1970). Registration of provar soybeans. *Crop Science*, *10*, 728. https://doi.org/10.2135/cropsci1970.0011183X001000060055x

Widener, S., Graef, G., Lipka, A. E., & Jarquin, D. (2021). An assessment of the factors influencing the prediction accuracy of genomic prediction models across multiple environments. *Frontiers in Genetics*, *12*, 689319. https://www.frontiersin.org/articles/10.3389/fgene.2021.689319

Wilcox, J. R., Schapaugh, Jr, W. T., Bernard, R. L., Fehr, W. R., & Niehaus, M. H. (1979). Genetic improvement of soybeans in the Midwest. *Crop Science*, *19*, 803–805. https://doi.org/10.1002/csc2.20820

Wilcox, J. R., & Shibles, R. M. (2001). Interrelationships among seed quality attributes in soybean. *Crop Science*, *41*, 11–14. https://doi.org/10.2135/cropsci2001.41111x

Wilson, R. F. (2004). Seed composition. In H. R. Boerma & J. E. Specht (Eds.), *Soybean: Improvement, production, and uses* (3rd ed., pp. 621–677). ASA-CSSA-SSSA.

Wilson, S., Blaschek, K., & Mejia, E. G. (2005). Allergenic proteins in soybean: Processing and reduction of P34 allergenicity. *Nutrition Reviews*, *63*, 47–58. https://doi.org/10.1111/j.1753-4887.2005.tb00121.x

Wu, N., Lu, Q., Wang, P., Zhang, Q., Zhang, J., Qu, J., & Wang, N. (2020). Construction and analysis of GmFAD2-1A and GmFAD2-2A soybean fatty acid desaturase mutants based on CRISPR/Cas9 technology. *International Journal of Molecular Sciences*, *21*(3), 1104. https://doi.org/10.3390/ijms21031104

Xavier, A., & Habier, D. (2022). A new approach fits multivariate genomic prediction models efficiently. *Genetics, Selection, Evolution.*, *54*, 45. https://doi.org/10.1186/s12711-022-00730-w

Xia, N., Yang, M., Zhao, J., Shao, Y., Shi, Y., Yan, W., Wang, X., Han, Y., & Wang, Z. (2019). Genome-wide association analysis of 1-octen-3-ol content related to soymilk off-flavor in soybean seed. *Crop and Pasture Science*, *70*, 133–139. https://doi.org/10.1071/CP18423

Xu, H., Guo, Y., Qiu, L., & Ran, Y. (2022). Progress in soybean genetic transformation over the last decade. *Frontiers in Plant Science*, *13*, 900318. https://doi.org/10.3389/fpls.2022.900318

Xu, H., Zhang, L., Zhang, K., & Ran, Y. (2020). Progresses, challenges, and prospects of genome editing in soybean (*Glycine max*). *Frontiers in Plant Science*, *11*, 571138. https://doi.org/10.3389/fpls.2020.571138

Yaklich, R. W., Vinyard, B., Camp, M., & Douglass, S. (2002). Analysis of seed protein and oil from soybean northern and southern region uniform tests. *Crop Science*, *42*, 1504–1515. https://doi.org/10.2135/cropsci2002.1504

Yang, A., & James, A. T. (2014). Influence of globulin subunit composition of soybean proteins on silken tofu quality. 1. Effect of growing location and 11SA4 and 7S alpha' deficiency. *Crop & Pasture Science*, *65*(3), 259–267. https://doi.org/10.1071/CP13398

Yang, A., Smyth, H., Chaliha, M., & James, A. (2016). Sensory quality of soymilk and tofu from soybeans lacking lipoxygenases. *Food Science & Nutrition*, *4*, 207–215. https://doi.org/10.1002/fsn3.274

Yang, W., Feng, H., Zhang, X., Zhang, J., Doonan, J. H., Batchelor, W. D., Xiong, L., & Yan, J. (2020). Crop phenomics and high-throughput phenotyping: Past decades, current challenges, and future perspectives. *Molecular Plant*, *13*(2), 187–214. https://doi.org/10.1016/j.molp.2020.01.008

Yoshino, M., Kanazawa, A., Tsutsumi, K.-I., Nakamura, I., Takahashi, K., & Shimamoto, Y. (2002). Structural variation around the gene encoding the α subunit of soybean β-conglycinin and correlation with the expression of the α subunit. *Breed Science*, *52*, 285–292. https://doi.org/10.1270/jsbbs.52.285

Yuan, W., Wu, Z., Zhang, Y. E., Yang, R., Wang, H., Kan, G., & Yu, D. (2021). Genome-wide association studies for sulfur-containing amino acids in soybean seeds. *Euphytica*, *217*, 155. https://doi.org/10.1007/s10681-021-02888-8

Zaidi, S. S.-A., Mahas, A., Vanderschuren, H., & Mahfouz, M. M. (2020). Engineering crops of the future: CRISPR approaches to develop climate-resilient and disease-resistant plants. *Genome Biology*, *21*, 289. https://doi.org/10.1186/s13059-020-02204-y

Zeng, A., Chen, P., Shi, A., Wang, D., Zhang, B., Orazaly, M., Florez-Palacios, L., Brye, K., Song, Q., & Cregan, P. (2014). Identification of quantitative trait loci for sucrose content in soybean seed. *Crop Science*, *54*(2), 554–564. https://doi.org/10.2135/cropsci2013.01.0036

Zeng, G., Li, D., Han, Y., Teng, W., Wang, J., Qiu, L., & Li, W. (2009). Identification of QTL underlying isoflavone contents in soybean seeds among multiple environments. *Theoretical and Applied Genetics*, *118*(8), 1455–1463. https://doi.org/10.1007/s00122-009-0994-5

Zhang, B., Chen, P., Shi, A., Hou, A., Ishibashi, T., & Wang, D. (2009). Putative quantitative trait loci associated with calcium content in soybean seed. *Journal of Heredity*, *100*, 263–269. https://doi.org/10.1093/jhered/esn096

Zhang, B., Tamura, M., Berger-Doyle, J., & Chen, P. (2008). Comparison of instrumental methods for measuring seed hardness of food-grade soybean. *Journal of Texture Studies*, *39*, 28–39. https://doi.org/10.1111/j.1745-4603.2007.00128.x

Zhang, F., Chen, C., Ge, H., Liu, J., Luo, Y., Liu, K., Chen, L., Xu, K., Zhang, Y., Tan, G., & Li, C. (2014). Efficient soybean regeneration and Agrobacterium-mediated transformation using a whole cotyledonary node as an explant. *Biotechnology and Applied Biochemistry*, *61*(5), 620–625. https://doi.org/10.1002/bab.1207

Zhang, H., Goettel, W., Song, Q., Jiang, H., Hu, Z., Wang, M. L., & An, Y. Q. C. (2020). Selection of GmSWEET39 for oil and protein improvement in soybean. *PLoS Genetics*, *16*(11), e1009114. https://doi.org/10.1371/journal.pgen.1009114

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

Zhang, K., Liu, S., Li, W., Liu, S., Li, X., Fang, Y., Zhang, J., Wang, Y., Xu, S., Zhang, J., Song, J., Qi, Z., Tian, X., Tian, Z., Li, W.-X., & Ning, H. (2018). Identification of QTNs controlling seed protein content in soybean using multi-locus genome-wide association studies. *Frontiers in Plant Science*, *9*, 1690. https://doi.org/10.3389/fpls.2018.01690

Zhang, L. X., Kyei-Boahen, S., Zhang, J., Zhang, M. H., Freeland, T. B., Watson, C. E., & Liu, X. (2007). Modifications of optimum adaptation zones for soybean maturity groups in the USA. *Crop Manage*, *6*(1), 1–11. https://doi.org/10.1094/CM-2007-0927-01-RS

Zhang, P., Burton, J. W., Upchurch, R. G., Whittle, E., Shanklin, J., & Dewey, R. E. (2008). Mutations in a Δ^9-Stearoyl-ACP-desaturase gene are associated with enhanced stearic acid levels in soybean Seeds. *Crop Science*, *48*(6), 2305–2313. https://doi.org/10.2135/cropsci2008.02.0084

Zhang, S., Du, H., Ma, Y., Li, H., Kan, G., & Yu, D. (2021). Linkage and association study discovered loci and candidate genes for glycinin and β-conglycinin in soybean (*Glycine max* L. Merr.). *Theoretical and Applied Genetics*, *134*, 1201–1215. https://doi.org/10.1007/s00122-021-03766-6

Zheng, L., Regenstein, J. M., Teng, F., & Li, Y. (2020). Tofu products: A review of their raw materials, processing conditions, and packaging. *Food Science and Food Safety*, *19*, 3683–3714. https://doi.org/10.1111/1541-4337.12640

Zhong, X., Hong, W., Shu, Y., Li, J., Liu, L., Chen, X., Islam, F., Zhou, W., & Tang, G. (2022). CRISPR/Cas9 mediated gene-editing of GmHdz4 transcription factor enhances drought tolerance in soybean (*Glycine max* [L.] Merr.). *Frontiers in Plant Science*, *13*, 988505. https://doi.org/10.3389/fpls.2022.988505

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: Singer, W. M., Lee, Y.-C., Shea, Z., Vieira, C. C., Lee, D., Li, X., Cunicelli, M., Kadam, S. S., Khan, M. A. W., Shannon, G., Mian, M. A. R., Nguyen, H. T., & Zhang, B. (2023). Soybean genetics, genomics, and breeding for improving nutritional value and reducing antinutritional traits in food and feed. *The Plant Genome*, *16,* e20415. https://doi.org/10.1002/tpg2.20415