SOLUBLE CARBOHYDRATES IN SOYBEAN - DISCOVERY OF NOVEL GENETIC RESOURCES AND EVALUATION OF KNOWN MUTATIONS

A Dissertation

presented to

the Faculty of the Graduate School

at the University of Missouri-Columbia

In Partial Fulfillment

of the Requirements for the Degree

Doctor of Philosophy

---

by

DONGHO LEE

Drs. James G. Shannon and Henry T. Nguyen, Dissertation Supervisors

Dr. Pengyin Chen (in memoriam)

MAY 2024

The undersigned, appointed by the Dean of the Graduate School, have examined the dissertation entitled:

SOLUBLE CARBOHYDRATES IN SOYBEAN - DISCOVERY OF NOVEL GENETIC RESOURCES AND EVALUATION OF KNOWN MUTATIONS

presented by Dongho Lee,

a candidate for the degree of Doctor of Philosophy,

and hereby certify that, in their opinion, it is worthy of acceptance.

Dr. James G. Shannon

Dr. Henry T. Nguyen

Dr. Tri D. Vuong

Dr. Andrew Scaboo

Dr. Dong Xu

DEDICATIONS

I love you, Lord, my strength.

The Lord is my rock, my fortress and my deliverer. My God is my rock, in whom I take refuge, my shield and the horn of my salvation, my stronghold. (Psalm 18:1-2)

Lord, thank you for always being with me on this journey. Your unconditional and everlasting love has always amazed me. I am dedicating this dissertation to you.

ACKNOWLEDGEMENTS

First and foremost, I would like to thank my family for their love, support, dedication, and sacrifice. Especially, I would like to express my gratitude to my wife, Bomi Kim. Your love is always special and encourages me when I am worn out. Without you, I wouldn't be able to complete this journey. We will continue to move forward for our dreams and be reliable supporters of each other for the rest of our lives.

I would like to express my sincere gratitude to Dr. Pengyin Chen for his warmest heart that allowed me to start this incredible journey in one of the best soybean breeding programs. Your guidance and dedication to helping me think, act, and live as a breeder were always grateful, and I miss it. Your exceptional enthusiasm for soybean breeding will never be forgotten. It will be forever fuel for me to live as a breeder.

In addition, I would like to thank the dissertation committee members, Dr. James G. Shannon, Dr. Henry T. Nguyen, Dr. Tri D. Vuong, Dr. Andrew Scaboo, and Dr. Dong Xu, for their guidance, advice, and support for my doctorate research. Whenever I felt lost, you were always available and supportive so that I could find a way. I am honored to work with all my committee members in my Ph.D.

My special thanks go to the soybean crew at the Fisher Delta Research, Extension, and Education Center. I sincerely appreciate your hospitality, friendship, and incredible support for me and my doctorate research. Every moment I spent with you guys was always

fun and memorable. For six years, I have received the kindest welcome and a sense of belonging. None of these would be possible without our soybean crew.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

Chapter 2: Novel genetic resources associated with sucrose and stachyose content through genome-wide association study in soybean [*Glycine max* (L.) Merr.]

ACADEMIC ABSTRACT

Soybean [*Glycine max* (L.) Merr.] is the leading protein meal source, primarily used to feed non-ruminant animals, such as poultry and swine. Additionally, soy products have garnered huge popularity as a high-quality plant-based protein source in vegetarian diets worldwide. Among soluble carbohydrates, sucrose positively impacts feed efficiency and the natural sweetness of soy products. In contrast, raffinose and stachyose (RFOs) create unpleasant intestinal conditions and reduce feed efficiency in non-ruminant animals. Therefore, this research focused on three main topics, including i) the identification of novel genetic resources for high sucrose and low stachyose content via genome-wide association study (GWAS), ii) the discovery of novel quantitative trait loci (QTL) and variants for high sucrose content via QTL mapping, and iii) the evaluation of environmental stability and genetic effect of soybean lines differing in allele combinations between raffinose synthase (RS) and D-myo-inositol-3-phosphate synthase 1 (MIPS1) mutations for soluble carbohydrates. Through GWAS, seven significant single nucleotide polymorphisms (SNPs) for sucrose content were identified across chromosomes (Chrs.) 2, 8, 12, 17, and 20, whereas 13 significant SNPs for stachyose content were identified across Chrs. 2, 5, 8, 9, 10, 13, 14, and 15. Among the candidate genes, *Glyma.08g361200* and *Glyma.17g258100* for sucrose content and *Glyma.05g025300* and *Glyma.13g077900* for stachyose content were highlighted based on functional annotations, protein-protein interactions, and RNA-seq Atlas database. For QTL mapping, a high sucrose germplasm, PI 506593, was selected from the previous GWAS analysis and used as a donor parent in two bi-parental mapping populations. The QTL mapping pinpointed a major QTL region on Chr. 8 at the physical interval of 40,597,410 – 42,861,364 bp. The whole-genome

sequencing (WGS) analysis using parental lines identified unique variants in 31 candidate genes within the major QTL region based on Wm82.a2.v1 and Wm82.a5.v1 reference genomes. Among those candidate genes, *Glyma.08g294300*, *Glyma.08g297000*, and *Glyma.08g300400* were closely associated with sucrose metabolism. Lastly, the environmental stability test and overall genetic effects of known mutant lines for high sucrose and low RFOs demonstrated that the *mips1* mutant lines had significantly higher sucrose and lower RFOs content than other lines with *rs2* and *rs3* mutations across the environments. The three different germination tests showed that the germination rates of mutant lines with *rs2W331*- and *mips1* mutations were either statistically similar or significantly greater than the check. Given the limited genetic resources and functional germplasms for desirable soluble carbohydrate profiles, this research provided valuable genetic information to maximize animal feed efficiency and market preference for soy products.

CHAPTER 1

Literature Review

The journey of soybean to the world

The history of modern cultivated soybean [*Glycine max* (L.) Merr.] began from wild soybean (*Glycine soja* Sieb. & Zucc.) in East Asia, although the precise information regarding regions and times is still ambiguous. Hymowitz (1970, 2004) proposed that soybean domestication first took place in the eastern half of north China during the Shang dynasty, approximately from 1700 to 1100 B.C. Until the first century A.D., soybean was spread out to different regions in China and peninsular Korea because of the population movement resulting from the degeneration of Chinese dynasties and the consolidation of new territories (Hymowitz, 2004; Wilson, 2008). Other studies supported another hypothesis that the domestication of modern cultivated soybean began 6,000 – 9,000 years ago (Carter et al., 2004; Kim et al., 2012). Recent archeological discoveries using molecular marker data suggested more precise regional origins of soybean, such as the Huanghe River (Yellow River), the Yangtze basin (Southern China), and northern and central China, considered the primary gene centers (Guo et al., 2010; Li et al., 2010; Wang et al., 2016). While the single-origin hypothesis of soybean domestication has been widely accepted, several studies using chloroplast or nuclear microsatellite markers in wild and cultivated soybeans have suggested the multiple-origin hypothesis in East Asia (Xu et al., 2002; Abe et al., 2003; Sedivy et al., 2017). Also, Fang et al. (2016) analyzed the chloroplast genomes of 302 wild, landraces, and modern soybeans, showing multiple maternal lines had been involved during soybean domestication. In addition, gene pools in chloroplast and nuclear genomes of Korean and Japanese soybeans were significantly different (Xu et al., 2002; Abe et al., 2003; Zhou et al., 2015). Moreover, the genetic diversity and population structure analysis showed that the *G. max* landraces from China

were genetically different from those from Korea and Japan (Li and Nelson, 2001; Lee et al., 2011). It suggested the possibility that *G. soja* had been initially grown in multiple origins, including China, Korea, and Japan, and domesticated independently (Xu et al., 2002). Despite all the scientific evidence based on archeological, molecular, and genomic studies, further studies will be required to clarify the history of soybean domestication.

From the 1st to 16th centuries, soybean has been continuously disseminated to other Asian countries, including Japan, Indonesia, the Philippines, Vietnam, Thailand, and Nepal (Hymowitz, 1990). In the late 16th century, soybean was first introduced in the Western hemisphere in Europe by missionaries and sailors, and they were planted in botanic gardens in France and England throughout the 17th and 18th centuries (Singh and Hymowitz, 1999). In 1765, Samuel Bowen, an English entrepreneur, farmer, and seaman, brought soybean seeds from China to the U.S. and cultivated the first soybean crop in the U.S. (Hymowitz and Harlan, 1983). Soybean was continuously brought from China to the U.S. by various individuals, such as scientists, seed dealers, merchants, and military expeditions (Singh and Hymowitz, 1999). Although the first cultivation of soybean in Latin America was estimated to start between 1565 and 1815, the first written reference about soybean was published by Gustavo D'Utra in 1882 (Shurtleff and Aoyagi, 2004). After several thousands of years of journey, soybean has become one of the most important legume crops, currently being planted on almost 130 million hectares across the world, and the global soybean production reached around 360 million metric tons (USDA, 2022; Soystats, 2023).

The economic impact of soybean

During the last decades, soybean has been established as one of the most important legume crops, providing high-quality protein and oil for human and animal consumption. The economic impact of the overall soybean industry has been significant in the U.S. (Vieira and Chen, 2021). In 2022, the U.S. soybean contributed over $124 billion to the U.S. economy, equivalent to almost 0.6% of the U.S. GDP (Gross Domestic Product) (LMC International, 2023). Furthermore, soybean is the most economically valuable crop in Brazil, which is the world's leading soybean producer. The economic value of soybean in Brazil was almost 350 billion BRL (equivalent to $71 billion) in 2022, which supported 1.4 million employees across the country (Czapp, 2023). Global soybean production is mainly consumed in three categories: 1) direct human food (20%), 2) animal feed (76%), and 3) industry usage (4%) (Ritchie, 2021). The largest portion of soybean production is consumed by animals, including poultry (37%) and swine (20%) (Ritchie, 2021). Recently, the number of companion pet owners seeking plant-based pet foods, such as soybean-based products (soy products), has been growing due to animal welfare, ethical, and moral concerns (Yoosefzadeh-Najafabadi et al., 2022). The pet industry contributes a significant portion of the U.S. economy, of which approximately 66% of U.S. households own a pet and spend almost $58.1 billion annually on pet food and treats. In addition, as the global aquaculture business has been expanding to meet the growing fish demands, soybean has been highlighted as a promising source of high-quality and low-cost protein feed for fish (He et al., 2020; Yang et al., 2022a). As human foods, soy products, including tofu and edamame, have recently garnered more attention from the global vegan population as a high-value protein substitute for animal meats (Qin et al., 2022). Soybean oil is a dominant vegetable oil in the U.S. market, accounting for almost 56% of the U.S. vegetable oil

consumption (SoyStats, 2023). Internationally, soybean oil ranked second place (29%) after palm oil (34%) in global vegetable oil consumption in 2022 (SoyStats, 2023). For industry usage, soybean oil has been highlighted in the biodiesel industry as a nontoxic and low-cost alternative due to its renewable and environmentally friendly traits (Uysal et al., 2017; Gebremariam and Marchetti, 2018). Also, there was a massive investment of $100 million from USDA for biodiesel infrastructure to support the sustainability of the biodiesel market (USDA, 2020). Currently, Brazil accounts for 51% of world soybean exports, followed by the U.S. (38%) and Argentina (3%), while China accounts for 70% of world soybean imports, followed by Mexico (3%) and Japan (3%) (SoyStats, 2023; TrendEconomy, 2023).

Seed compositional traits for soymeal

The unique quality of soybean seed composition has made soybean an economically important and most-produced legume crop globally (Semba et al., 2021; Vieira and Chen, 2021). Soybean seeds are typically composed of three main components: protein (40%), oil (20%), and carbohydrates (35%) on a dry-weight basis (Hsu et al., 1973; Liu, 1997). To maximize its utility and value, soybean is usually processed to separate oil from other components, which are called soymeal, mainly protein and carbohydrates (Singer et al., 2023). Soymeal is the leading source of high-quality protein supplementary with all nine essential amino acids (histidine, isoleucine, leucine, lysine, methionine, phenylalanine, threonine, tryptophan, and valine) for animal feeds (Singer et al., 2023). More importantly, soymeal meets three fundamental requirements in animal feeding programs, which are 1) having one or more essential nutrients, 2) being available to meet

the demand of regular usage on a large scale, and 3) being cost-effective to use (Dozier and Hess, 2011).

Another important soymeal component is soluble carbohydrates, which are responsible for metabolizable energy (ME) efficiency in animal feeds. Soybean seeds typically contain up to 15% soluble carbohydrates on a dry-weight basis (Feng et al., 2005), which have three main components: sucrose, raffinose, and stachyose. Sucrose is the only nutritionally beneficial component readily digestible to produce ME, while raffinose and stachyose (raffinose family of oligosaccharides, RFOs) are considered antinutrients reducing ME efficiency in animal feeds (Jo et al., 2018; Wang et al., 2023). Since the ME of sucrose (3,900 kcal/kg) is significantly greater than that of starch (2,918 – 3,396 kcal/kg), the U.S. animal growers recommend optimized feeding formulations with higher sucrose concentrations (John, 2008; Ostezan et al., 2023). However, RFOs are indigestible in monogastric animals, including poultry, swine, dogs, and cats, due to the lack of α-galactosidase in their gastric systems to break down the α-1,6-glycosidic bonds of RFOs. Indigested RFOs pass through the upper gut and are fermented by anaerobic microbes in the lower gut (Le et al., 2020). It produces unwanted gastric discomforts, such as diarrhea and flatulence, resulting in a significant loss of ME efficiency (Coon et al., 1990; Le et al., 2020).

Soybean-based foods (Soy foods)

Although soy foods have been widely consumed in Asia for centuries, they became known relatively late in Western countries (Messina and Messina, 2010). In recent decades, soy foods have rapidly become the most popular plant-based protein source in Western

diets for ethical, ecological, and health purposes (Messina and Messina, 2010; Rizzo and Baroni, 2018). Also, vegetarians value the ease of incorporating soy foods into their plant-based diets (Messina and Messina, 2010). The main soy foods commercialized in the markets include tofu, soy milk, miso, natto, soy flour, and edamame. Tofu, made from soybean protein, is gelatinous and readily digestible, improving nutrient absorption in the human body (Huang et al., 2022). The high global demand for tofu products has recently led to efforts to improve its taste (Joo and Cavender, 2020; Shi et al., 2020). Soy milk is produced through several aqueous extraction processes, including soaking, mechanical grinding, extracting the insoluble fraction, and pasteurizing the milk solution (Davy and Vuong, 2022). Soy milk contains fats, carbohydrates, proteins, vitamins, minerals, isoflavones, and soy saponins. It is widely consumed to replace cow's milk due to its fewer allergic reactions and additional nutritional benefits (Qin et al., 2022; Storz et al., 2023). Miso and natto (known as 'Chungkookjang' in Korea) are traditional soy foods in Japan, made by fermenting whole beans for around six months and 2 – 4 days, respectively (Lee et al., 2019a; Allwood et al., 2021). The fermentation efficiently enhances the nutritional value of soybean by increasing the number of bioactive peptides and food-grade hydrolyzed proteins, resulting in various health benefits, such as strengthening immune systems, effects of anti-cancer, anti-inflammatory, anti-genotoxic, anti-oxidant, and fibrinolytic, lowering blood pressure (Lee et al., 2019a). Due to the health benefits of soybean, soy flour has been added to wheat flour to improve bread quality (Teleky et al., 2020). Vegetable soybean, called edamame, is a food-grade soybean harvested when seeds are fully developed while the pods are still green (Moseley et al., 2021; Khan et al., 2023). Edamame is a versatile high-protein ingredient mainly used in soups, stews, salads, and

snacks, and it is the second most consumed soy food in the U.S. after soy milk (Yu et al., 2021; Khan et al., 2023). Edamame typically contains a higher level of sugar, protein, minerals, and vitamins with a larger seed size than grain soybean (Khan et al., 2023). Among other traits, sucrose content contributes to the natural sweetness of soy foods. Thus, soybean with elevated sucrose content is desirable primarily for various soy foods (Rosset et al., 2012; Lee et al., 2023; Wang et al., 2023).

Biosynthesis of sucrose

In photosynthesis, plants produce triose phosphates through carbon fixation in the Calvin cycle in the chloroplast. Then, two triose phosphate molecules are used to produce fructose 1,6-bisphosphate (F1,6BP) by aldolase in the cytosol. With the catalyzation of hexose phosphatase, F1,6BP is metabolized to fructose 6-phosphate (F6P), subsequently producing glucose 6-phosphate (G6P), which is primarily used for UDP-glucose formation (Cheikh and Brenner, 1992; Ruan, 2014; Stein and Granot, 2019). Two essential substrates, F6P and UDP-glucose, are used in sucrose-phosphate production, and sucrose-phosphate is subsequently converted into sucrose (disaccharide) by sucrose phosphate phosphatase (Ruan, 2014; Stein and Granot, 2019). Sucrose, the final product of photosynthesis, is then loaded into the phloem and transported from the leaf (source cells) to non-photosynthetic tissues (sink cells), such as roots, modified leaves, and seeds (Patil et al., 2015).

Biosynthesis of galactinol and myo-inositol

Galactinol is an important intermediate used as a donor of the galactosyl group, which is leading a carbon chain elongation in the biosynthesis of RFOs. Galactinol is

synthesized from uridine 5'-diphosphate-galactose (UDP-galactose) and myo-inositol by the catalyzation of galactinol synthase (UDP-galactose:inositol galactosyltransferase) (Saravitz et al., 1987). Myo-inositol is an essential substrate in galactinol biosynthesis and plays a key role in plant growth development, cell wall biosynthesis, communication between cells, and plant hormone transportation (Irvine and Schell, 2001). The sole synthetic source of myo-inositol backbone (D-myo-inositol-3-phosphate) is converted from G6P by the catalyzation of D-myo-inositol-3-phosphate synthase (MIPS) (Loewus and Murthy, 2000; Redekar et al., 2020).

Biosynthesis of raffinose family of oligosaccharides, RFOs

The biosynthesis of raffinose (trisaccharide) uses sucrose as a main substrate and is catalyzed by raffinose synthase to produce raffinose. Raffinose synthase (RS) executes an initial chain elongation by transferring one galactosyl group of galactinol to sucrose (Peterbauer and Richter, 2001). Likewise, the second chain elongation is executed by stachyose synthase by transferring another galactosyl group to raffinose to produce stachyose (tetrasaccharide) (Peterbauer and Richter, 2001). The galactosyl group is connected by α-1,6-glycosidic bonds during the chain elongation (Kumar et al., 2010).

Environmental impacts on soluble carbohydrates

Previous studies elucidated the effects of growing conditions on seed composition traits, including carbohydrates, protein, oil, and fatty acids in soybean (Nian et al., 1996; Piper and Boote, 1999; Thomas et al., 2003; Pipolo et al., 2004; Hou et al., 2006; Lee et al., 2009; Kumar et al., 2010; Bilyeu and Wiebold, 2016; Jo et al., 2018). Under various

environmental factors, vegetative growth, flowering timing, reproductive stages, and seed maturation can be influenced by gene expression changes and enzyme activity during overall plant growth (Vicente-Carbajosa and Carbonero, 2005; Cho et al., 2017). For soluble carbohydrates, two factors, temperature and soil moisture content during the pod-filling stages, were reported to be significantly associated with the accumulation of soluble carbohydrates (Bellaloui, 2012; Bilyeu and Wiebold, 2016; Jo et al., 2018; Wijewardana et al., 2019; Du et al., 2020; Wang et al., 2021). The temperature during the pod-filling stages was negatively associated with soybean sucrose content (Bilyeu and Wiebold, 2016; Jo et al., 2018). However, the association between temperature and RFOs is still uncertain. Kumar et al. (2010) and Jo et al. (2018) reported that RFOs content was genotype-dependent across environments, while Bilyeu and Wiebold (2016) observed that the temperatures during pod-filling stages positively correlated with RFOs content. Also, several studies showed a soil moisture deficit increased sucrose and raffinose content and decreased stachyose content in soybean (Bellaloui, 2012; Wijewardana et al., 2019; Du et al., 2020; Wang et al., 2023).

Functional mutations for desirable soluble carbohydrate profiles

The desirable profile of increased sucrose and reduced RFOs content was observed in the plant introduction (PI) 200508 in the USDA Soybean Germplasm Collection (Kerr and Sebastian, 2000). PI 200508 was found to contain a three-base pair deletion within the raffinose synthase 2 coding regions (RS2, *Glyma.06g179200*, Wm82.a2.v1), resulting in one amino acid deletion (tryptophan, W) at a highly conserved position 331 (*rs2W331-*; Kerr and Sebastian, 2000; Dierking and Bilyeu, 2008; Skoneczka et al., 2009). A missense

mutation in the RS2 coding sequence was later found in an ethyl-methanesulfonate (EMS) mutant line using reverse genetics by Targeting Induced Local Lesions IN Genomes (TILLING) (Dierking and Bilyeu, 2009). This mutation changed one amino acid from threonine (T) to isoleucine (I) at position 107 (*rs2T107I*) (Dierking and Bilyeu, 2009). Later, Hagely et al. (2013) found that the ultra-low RFOs content was achieved with the presence of synonymous RS3 (*Glyma.05g003900*, Wm82.a2.v1) variant alleles (*rs3snp5*/*rs3snp6*) and RS2 mutations together. However, only variant alleles of RS3 had no significant effects on the soluble carbohydrate profile (Jo et al., 2018). With chemical mutagenesis, Sebastian et al. (2000) developed a mutant line LR33 carrying a single base pair substitution in a conserved region of D-myo-inositol-3-phosphate synthase 1 (MIPS1) gene (*Glyma.11g238800*, Wm82.a2.v1). Although MIPS1 mutation significantly increased sucrose content and reduced RFOs, MIPS1 mutant lines showed detrimental effects on agronomic traits, including germination rate, during the field experiments (Meis et al., 2003; Chappell et al., 2006). Years later, Saghai Maroof and Buss (2008) developed a soybean line V99-5089, which was derived from a cross of V71-370 × PI 87013. The line contained a natural variant of MIPS1 and showed a desirable soluble carbohydrate profile (> 11% sucrose and < 1% RFOs). Moreover, the nucleic acid sequence analysis revealed that the MIPS1 mutation in V99-5089 differed from that in LR33. To date, three genes, *Glyma.05g003900* (RS3), *Glyma.06g179200* (RS2), and *Glyma.11g238800* (MIPS1), have been widely used in marker-assisted selection breeding to achieve desirable soluble carbohydrate profiles.

Previous QTL mapping studies for soluble carbohydrates

Quantitative trait loci (QTL) mapping has been employed to identify genomic loci governing complex traits using segregating bi-parental populations. This approach requires two distinct parents for a trait of interest to create F_2, backcross (BC), recombinant inbred lines (RILs), and near-isogenic lines (NILs) population. Recently, multi-parent populations called multi-parent advanced generation intercross (MAGIC), chromosome segment substitution lines (CSSLs), and nested-association mapping populations have been used for QTL mapping (Bandillo et al., 2013; Ogawa et al., 2016; Bessho-Uehara et al., 2017; Frangoso et al., 2017). A genetic linkage map is constructed based on the molecular markers tightly linked to QTL to localize genomic regions associated with the traits of interest. In soybean, QTL mapping has been widely used to identify genomic regions that control complex quantitative traits, such as seed compositions (Rani et al., 2023). Compared to protein and oil, soluble carbohydrates have garnered less attention from soybean breeders. To date, bi-parental mapping studies have reported 248 QTL related to protein and 322 QTL related to oil. However, only 52 QTL related to soluble carbohydrates have been reported, including 37 QTL for sucrose and 15 QTL for RFOs. (Soybase.org). Maughan et al. (2000) reported 17 sucrose-related QTL on chromosomes (Chrs.) 5, 7, 8, 13, 15, 19, and 20 using 149 $F_{2:4}$ RILs from an interspecific cross between a large-seeded breeding line and a *G. soja* line. Kim et al. (2005) identified QTL related to sucrose and RFOs on Chrs. 2, 11, and 19, with common QTL on Chr. 2 and 19. Later, Kim et al. (2006) identified QTL on Chrs. 2, 6, 12, 14, 15, 16, and 19 using two bi-parental populations sharing the same high sucrose donor line (Keunolkong), of which five QTL for sucrose and RFOs on Chrs. 12, 14, 15, and 16 were located in the same genomic regions. A major QTL for sucrose and RFOs was identified on Chr. 11 using a population derived from a cross

between a large-seeded breeding line (V71-370) and *G. max* line (PI 87013) (Saghai Maroof and Buss, 2008). Skoneczka et al. (2009) used two F_2-derived bi-parental populations (crosses PI 87013 × PI 200508 and PI 243545 × PI 200508) to identify another major QTL on Chr. 6 for increased sucrose and reduced RFOs. Later, Wang et al. (2014) identified three sucrose- and four RFOs-related QTL on Chrs. 7, 11, 12, and 20, of which a QTL on Chr. 11 was located on the same genomic region with the QTL that Saghai Maroof and Buss (2008) identified. Zeng et al. (2014) first used high-density SNP markers and identified three sucrose-related QTL on Chrs. 5, 9, and 16, accounting for 46, 10, and 8% phenotypic variations, respectively. Akond et al. (2015) identified three and 11 QTL for sucrose and RFOs, respectively, across Chrs. 1, 3, 6, 9, 12, 14, 15, and 16 using 92 $F_{5:7}$ RILs from a cross of MD96-5722 × Spencer. To date, many bi-parental mapping studies using different donor parents and larger marker sets have been conducted to identify QTL related to soluble carbohydrates (Lee et al., 2016; Patil et al., 2018; Salari et al., 2021; Knizia et al., 2023; Liu et al., 2023).

Previous genome-wide association studies (GWAS) for soluble carbohydrates

Genome-wide association study (GWAS) is an alternative strategy to QTL mapping to reveal the genetic architecture underlying complex quantitative traits of interest in plant breeding, including seed compositions, disease resistance, and yield traits. GWAS requires diverse plant germplasms from different origins and maturity groups (MG) with a large set of genetic marker data (Korte and Farlow, 2013). The basic approach of GWAS is to discover an association between genetic markers and a trait of interest across diverse germplasm accessions using statistical methodologies (Korte and Farlow, 2013). To date,

GWAS has been successfully applied in soybean research to discover and characterize important traits, including seed protein and oil (Hwang et al., 2014; Li et al., 2019; Zhang et al., 2021), amino acids (Lee et al., 2019b; Yuan et al., 2021), fatty acids (Liu et al., 2020; Sung et al., 2021), disease resistance (Vuong et al., 2015; Ravelombola et al., 2020; Vieira et al., 2022; McDonald et al., 2023), abiotic stress tolerance (Kaler et al., 2017; Wu et al., 2019; Saleem et al., 2022), agronomic traits (Ayalew et al., 2022; Cao et al., 2022; Yang et al., 2022b), and root system (Seck et al., 2020; Rathore et al., 2022; Kim et al., 2023). GWAS for sucrose and RFOs in soybean is relatively new compared to the QTL mapping study, and only a few studies have implemented GWAS for soluble carbohydrates. Vaughn et al. (2014) conducted the first GWAS using two diverse panels to elucidate the genetic architecture of soluble carbohydrates. This study identified three sucrose- and two stachyose-associated SNP markers on Chrs. 5, 9, 15, 18, and 20 using Near Infrared Spectroscopy (NIRS) (Vaughn et al., 2014). Sui et al. (2020) used 178 soybean germplasms, including landraces and cultivars, as a diverse panel to conduct GWAS with a compressed mixed linear model (CMLM) in GAPIT (Lipka et al., 2012). This study identified five quantitative trait nucleotides (QTNs) associated with sucrose content across Chrs. 4, 6, 7, 11, and 12. Among 60 candidate genes located near five QTNs, ten genes were closely associated with sucrose content based on the gene-based association: one gene with known function and nine novel genes regarding sucrose metabolism (Sui et al., 2020). Ficht et al. (2022) later conducted GWAS using 266 historical, breeding, and elite materials as a panel to improve the knowledge of QTL associated with sucrose content in soybean. GWAS analyses were conducted using a fixed and random model circulating probability unification (FarmCPU; Liu et al., 2016) in the rMVP package (Yin et al., 2020) in R

software (Ficht et al., 2022). This study identified seven significant QTL on Chrs. 1, 6, 8, 9, 10, 13, and 14, of which a major gene responsible for sucrose synthase was found on Chr. 6 (Ficht et al., 2022). In 2023, three comprehensive GWASs for soluble carbohydrates were conducted by Hu et al. (2023), Riaz et al. (2023), and Wang et al. (2023). Hu et al. (2023) employed a general linear model (GLM) and mixed linear model (MLM) for GWAS using 323 PI lines of early maturity groups 0 and 00 to identify significant SNPs for fructose, glucose, sucrose, raffinose, stachyose, and total soluble carbohydrate content. With SoySNP50K BeadChips, Hu et al. (2023) identified 17 sucrose-related QTL (Chrs. 2, 3, 6, 7, 8, 9, 11, 12, 15, 18, and 20), 11 raffinose-related QTL (Chrs. 1, 2, 4, 5, 6, 11, 14, and 15), and 11 stachyose-related QTL (Chrs. 1, 2, 3, 6, 7, 9, 12, 14, and 18). A total of five candidate genes for sucrose (two genes), raffinose (two genes), and stachyose (one gene) were identified based on further analysis, including gene ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) (Hu et al., 2023). Riaz et al. (2023) used a diverse panel consisting of 473 accessions originating from 11 countries for GWAS. This study initially identified four stable and novel SNPs on Chrs. 3, 6, 9, and 14 using three GWAS models: GLM, FarmCPU, Bayesian-information, and linkage-disequilibrium iteratively nested keyway (BLINK; Huang et al., 2019). Twenty-three putative genes within the 20 Kbp flanking genomic regions of those stable SNPs showed high gene expression levels, of which the one on Chr. 6 had the highest expression in pod shells (Riaz et al., 2023). Wang et al. (2023) conducted GWAS using 189 edamame accessions. The accessions were harvested at R6-R7 growth stages (Fehr, 1971) and phenotyped by high-performance liquid chromatography (HPLC). Using MLM, a total of 45 SNPs and 41

candidate genes (four genes with known effects and 37 novel genes) were found to be associated with sucrose content in edamame (Wang et al., 2023).

REFERENCE

Abe, J., Xu, D., Suzuki, Y., Kanazawa, A., and Shimamoto, Y. (2003). Soybean germplasm pools in Asia revealed by nuclear SSRs. *Theoretical and Applied Genetics*, 106, 445-453. https://doi.org/10.1007/s00122-002-1073-3

Akond, M., Liu, S., Kantartzi, S.K., Meksem, K., Bellaloui, N., Lightfoot, D.A., et al. (2015). Quantitative trait loci underlying seed sugars content in "MD96-5722" by "Spencer" Recombinant inbred line population of soybean. *Food and Nutrition Sciences*, 6, 964–973. https://doi.org/10.4236/fns.2015.611100

Allwood, J.G., Wakeling, L.T., and Bean, D.C. (2021). Fermentation and the microbial community of Japanese koji and miso: A review. *Journal of Food Science*, 86(6), 2194-2207. https://doi.org/10.1111/1750-3841.15773

Ayalew, H., Schapaugh, W., Vuong, T., and Nguyen, H.T. (2022). Genome-wide association analysis identified consistent QTL for seed yield in a soybean diversity panel tested across multiple environments. *The Plant Genome*, 15(4), e20268. https://doi.org/10.1002/tpg2.20268

Bandillo, N., Raghavan, C., Muyco, P.A., Sevilla, M.A.L., Lobina, I.T., Dilla-Ermita, C.J., et al. (2013). Multi-parent advanced generation inter-cross (MAGIC) populations in rice: progress and potential for genetics research and breeding. *Rice*, 6, 11. https://doi.org/10.1186/1939-8433-6-11

Bellaloui, N. (2012). Soybean Seed Phenol, Lignin, and Isoflavones and Sugars Composition Altered by Foliar Boron Application in Soybean under Water Stress. *Food and Nutrition Sciences*, 3(4), 18507. https://doi.org/10.4236/fns.2012.34080

Bessho-Uehara, K., Furuta, T., Masuda, K., Yamada, S., Angeles-Shim, R.B., Ashikari, M., et al. (2017). Construction of rice chromosome segment substitution lines harboring *Oryza barthii* genome and evaluation of yield-related traits. *Breeding Science*, 67(4), 408-415. https://doi.org/10.1270/jsbbs.17022

Bilyeu, K., and 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

Cao, Y., Jia, S., Chen, L., Zeng, S., Zhao, T., and Karikari, B. (2022). Identification of major genomic regions for soybean seed weight by genome-wide association study. *Molecular Breeding*, 42, 38. https://doi.org/10.1007/s11032-022-01310-y

Carter, T.E, Hymowitz, T., and Nelson, R.L. (2004). Biogeography, local adaption, Vavilov, and genetic diversity in soybean. In D. Werner (Ed.), *Biological resources and migration* (pp. 47-59). Springer Berlin Heidelberg.

Chappell, A.S., Scaboo, A.M., Wu, X., Nguyen, H., Pantalone, V.R., and Bilyeu, K.D. (2006). Characterization of the MIPS gene family in Glycine max. *Plant Breeding*, 125(5), 493–500. https://doi.org/10.1111/j.1439-0523.2006.01264.x

Cheikh, N., and Brenner, M.L. (1992). Regulation of key enzymes of sucrose biosynthesis in soybean leaves. *Plant Physiology*, 100(3), 1230-1237. https://doi.org/10.1104/pp.100.3.1230

Cho, L.H., Yoon, J., and An, G. (2017). The control of flowering time by environmental factors. *The Plant Journal*, 90(4), 708-719. https://doi.org/10.1111/tpj.13461

Czapp (2023). Soybean in Brazil. Available at: *https://www.czapp.com/analyst-insights/czapp-explains-soybean-in-brazil/*. Accessed 17 December 2023.

Davy, P., and Vuong, Q.V. (2022). Soy Milk By-product: Its Composition and Utilisation. *Food Reviews International*, 38, 147-169. https://doi.org/10.1080/87559129.2020.1855191

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

Dierking, E.C., and 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

Dozier, W.A., and Hess, J.B. (2011). Soybean meal quality and analytical techniques. In H. El-Shemy (Ed.), *Soybean and Nutrition* (pp.111-124). IntechOpen.

Du, Y., Zhao, Q., Chen, L., Yao, X., Zhang, W., Zhang, B., et al. (2020). Effect of drought stress on sugar metabolism in leaves and roots of soybean seedlings. *Plant Physiology and Biochemistry*, 146, 1-12. https://doi.org/10.1016/j.plaphy.2019.11.003

Fang, C., Ma, Y., Yuan, L., Wang, Z., Yang, R., Zhou, Z., et al. (2016). Chloroplast DNA underwent independent selection from nuclear genes during soybean domestication and improvement. *Journal of Genetics and Genomics*, 43(4), 217-221. https://doi.org/10.1016/j.jgg.2016.01.005

Ficht, A., Bruce, R., Torkamaneh, D., Grainger, C.M., Eskandari, M., and 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

Fehr, W.R., Caviness, C.E., Burmood, D.T., and 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

Frangoso, C.A., Moreno, M., Wang, Z., Heffelfinger, C., Arbelaez, L.J., Aguirre, J.A., et al. (2017). Genetic Architecture of a Rice Nested Association Mapping Population. *Genes, Genomes, Genetics*, 7(6), 1913-1926. https://doi.org/10.1534/g3.117.041608

Gebremariam, S.N., and Marchetti, J.M. (2018). Economics of biodiesel production: Review. *Energy Conversion and Management*, 168, 74-84. https://doi.org/10.1016/j.enconman.2018.05.002

Guo, J., Wang, Y., Song, C., Zhou, J., Qiu, L., Huang, H., et al. (2010). A single origin and moderate bottleneck during domestication of soybean (*Glycine max*): implications from microsatellites and nucleotide sequences. *Annals of Botany*, 106(3), 505–514. https://doi.org/10.1093/aob/mcq125

Hagely, K.B., Palmquist, D., and Bilyeu, K. (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

He, M., Yu, Y., Li, X., Poolsawat, L., Yang, P., Bian, Y., et al. (2020). An evaluation of replacing fish meal with fermented soybean meal in the diets of largemouth bass (*Micropterus salmoides*): Growth, nutrition utilization, and intestinal histology. *Aquaculture Research*, 51(10), 4302-4314. https://doi.org/10.1111/are.14774

Hou, G., Ablett, G.R., Pauls, P., and Rajcan, I. (2006). Environmental effects on fatty acid levels in soybean seed oil. *Journal of the American Oil Chemists' Society*, 83, 759-763. https://doi.org/10.1007/s11746-006-5011-4

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

Hu, L., Wang, X., Zhang, J., Florez-Palacios, L., Song, Q., and Jiang, G.L. (2023). Genome-Wide Detection of Quantitative Trait Loci and Prediction of Candidate Genes for Seed Sugar Composition in Early Mature Soybean. *International Journal of Molecular Science*, 24(4), 3167. https://doi.org/10.3390/ijms24043167

Huang, M., Liu, X., Zhou, Y., Summers, R.M., and Zhang, Z. (2019). BLINK: a package for the next level of genome-wide association studies with both individuals and markers in the millions. *Gigascience*, 8(2), giy154. https://doi.org/10.1093/gigascience/giy154

Huang, Z., Liu, H., Zhao, L., He, W., Zhou, X., Chen, H., et al. (2022). Evaluating the effect of different processing methods on fermented soybean whey-based tofu quality, nutrition, and flavour. *LWT-Food science and technology*, 158, 113139. https://doi.org/10.1016/j.lwt.2022.113139

Hwang, E.Y., Song, Q., Jia, G., Specht, J.E., Hyten, D.L., Costa, J., et al. (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. (1970). On the domestication of the soybean. *Economic Botany*, 24, 408-421. https://doi.org/10.1007/BF02860745

Hymowitz, T. (1990). Soybean: the success story. In: J. Janick and J.E. Simon (Eds.), *Advances in new crops* (pp.159-163). Timber Press.

Hymowitz, T. (2004). Speciation and cytogenetics. In: H.R. Boerma and J.E. Specht (Eds.), *Soybeans: Improvement, production, and uses* (pp. 97-136). American Society of Agronomy.

Hymowitz, T. (2008). The history of the soybean. In *Soybeans* (pp. 1-31). AOCS Press, Elsevier. https://doi.org/10.1016/B978-1-893997-64-6.50004-4

Hymowitz, T., and Harlan, J.R. (1983). Introduction of soybean to North America by Samuel Bowen in 1765. *Economic Botany*, 37(4), 371-379. https://doi.org/10.1007/BF02904196

Irvine, R.F., and Schell, M.J. (2001). Back in the water: the return of the inositol phosphates. *Nature Review Molecular Cell Biology*, 2, 327-338. https://doi.org/10.1038/35073015

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

John, P.G.K. (2008). Sugar syrup: the new energy feed for poultry. *World Poultry Science*, 24, 12-13. Available at: https://edepot.wur.nl/9105. Accessed 11 January 2024.

Joo, K.H., and Cavender, G.A. (2020). Investigation of tofu products coagulated with trimagnesium citrate as a novel alternative to nigari and gypsum: Comparison of physical properties and consumer preference. *LWT-Food Science and Technology*, 118, 108819. https://doi.org/10.1016/j.lwt.2019.108819

Kaler, A.S., Ray, J.D., Schapaugh, W.T., King, C.A., and Purcell, L.C. (2017). Genome-wide association mapping of canopy wilting in diverse soybean genotypes. *Theoretical and Applied Genetics*, 130, 2203-2217. https://doi.org/10.1007/s00122-017-2951-z

Kerr, P.S., and Sebastian, S.A. (2000). Soybean products with improved carbohydrate composition and soybean plants. United States Patent 6,147,193.

Khan, F.A., Amir, M., Narayan, S., Dar, Z.M., and Khan, M.H. (2023). Vegetable soybean (Edamame): A potential area of research – A review. *SKUAST Journal of Research*, 25(3), 376-386. https://doi.org/10.5958/2349-297X.2023.00051.X

Kim, H.K., Kang, S.T., Cho, J.H., Choung, M.G., and 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., and 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.Y., Van, K., Kang, Y.J., Kim, K.H., and Lee, S.H. (2012). Tracing soybean domestication history: from nucleotide to genome. *Breeding Science*, 61, 445-452. https://doi.org/10.1270/jsbbs.61.445

Kim, S.H., Tayade, R., Kang, B.H., Hahn, B.S., Ha, B.K., and Kim, Y.H. (2023). Genome-Wide Association Studies of Seven Root Traits in Soybean (*Glycine max* L.) Landraces. *International Journal of Molecular Sciences*, 24(1), 873. https://doi.org/10.3390/ijms24010873

Knizia, D., Bellaloui, N., Yuan, J., Lakhssassi, N., Anil, E., Vuong, T., et al. (2023). Quantitative Trait Loci and Candidate Genes That Control Seed Sugars Contents in the Soybean 'Forrest' by 'Williams 82' Recombinant Inbred Line Population. *Plants*, 12(19), 3498. https://doi.org/10.3390/plants12193498

Korte, A., and Farlow, A. (2013). The advantages and limitations of trait analysis with GWAS: a review. *Plant Methods*, 9, 29. https://doi.org/10.1186/1746-4811-9-29

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

Le, H., Nguyen, N.H., Ta, D.T., Le, T.N.T., Bui, T.P., Le, N.T., et al. (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, D., Kulkarni, K.P., Kim, B., Seok, Y.M., Song, J.T., and Lee, J. (2019a). Comparative assessment of quality characteristics of Chungkookjang made from soybean seeds differing in oleic acid concentration. *Journal of Functional Foods*, 52, 529-536. https://doi.org/10.1016/j.jff.2018.10.016

Lee, D., Lara, L., Moseley, D., Vuong, T.D., Shannon, G., Xu, D., et al. (2023). Novel genetic resources associated with sucrose and stachyose content through genome-wide association study in soybean (*Glycine max* (L.) Merr.). *Frontiers in Plant Science*, 14, 1294659. https://doi.org/10.3389/fpls.2023.1294659

Lee, J.D., Woolard, M., Sleper, D.A., Smith, J.R., Pantalone, V.R., Nyinyi, C.N., et al. (2009). Environmental effects on oleic acid in soybean oil of plant introductions with elevated oleic concentration. *Crop Science*, 49(5), 1762-1768. https://doi.org/10.2135/cropsci2008.11.0663

Lee, J.D., Vuong, T.D., Moon, H., Yu, J.K., Nelson, R.L., and Nguyen, H.T. (2011). Genetic diversity and population structure of Korean and Chinese soybean [*Glycine max* (L.) Merr.]. *Crop Science*, 51(3), 1080-1088. https://doi.org/10.2135/cropsci2010.07.0420

Lee, J.S., Kim, S.M., and Kang, S. (2016). Fine mapping of quantitative trait loci for sucrose and oligosaccharide contents in soybean [*Glycine max* (L.) Merr.] using 180 K Axiom SoyaSNP genotyping platform. *Euphytica*, 208, 195-203. https://doi.org/10.1007/s10681-015-1622-x

Lee, S., Van, K., Sung, M., Nelson, R., LaMantia, J., McHale, L.K., et al. (2019b). 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, 1639-1659. https://doi.org/10.1007/s00122-019-03304-5

Li, Y.H., Li, W., Zhang, C., Yang, L., Chang, R.Z., Gaut, B.S., et al. (2010). Genetic diversity in domesticated soybean (*Glycine max*) and its wild progenitor (*Glycine*

soja) for simple sequence repeat and single-nucleotide polymorphism loci. *New Phytologist*, 188(1), 242–253. https://doi.org/10.1111/j.1469-8137.2010.03344.x

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

Li, Z., and Nelson, R.L. (2001). Genetic diversity among soybean accessions from three countries measured by RAPDs. *Crop Science*, 41(4), 1337-1347. https://doi.org/10.2135/cropsci2001.4141337x

Lipka, A.E., Tian, F., Wang, Q., Peiffer, J., Li, M., Bradbury, P.J., et al. (2012). GAPIT: Genome association and prediction integrated tool. *Bioinformatics*, 28, 2397-2399. https://doi.org/10.1093/bioinformatics/bts444

Liu, C., Chen, H., Yu, Q., Gu, H., Li, Y., Tu, B., et al. (2023). Identification of quantitative trait loci and candidate genes for seed sucrose and soluble sugar concentrations in soybean. *Crop Science*, 63, 2976-2992. https://doi.org/10.1002/csc2.21080

Liu, K. (1997). Chemistry and nutritional value of soybean components. In *Soybeans* (pp. 25-113). Springer.

Liu, X., Qin, D., Piersanti, A., Zhang, Q., Miceil, C., and Wang, P. (2020). Genome-wide association study identifies candidate genes related to oleic acid content in soybean seeds. *BMC Plant Biology*, 20, 399. https://doi.org/10.1186/s12870-020-02607-w

Liu, X.L., Huang, M., Fan, B., Buckler, E.S., and Zhang, Z. (2016). Iterative usage of fixed and random effect models for powerful and efficient genome-wide association studies. *PLOS Genetics*, 12(3), e1005767. https://doi.org/10.1371/journal.pgen.1005767

LMC International (2023). The Economic Impact of U.S. Soybeans and End Products on the U.S. Economy – 2023 Update. Oxford, U.K. Available at: https://www.nopa.org/wp-content/uploads/2023/08/0LMC_SoyEconStudy_Aug2023.pdf. Accessed 12 January 2024.

Loewus, F.A., and Murthy, P.P.N. (2000). Myo-inositol metabolism in plants. *Plant Science*, 150(1), 1-19. https://doi.org/10.1016/S0168-9452(99)00150-8

Maughan, P.J., Saghai Maroof, M.A., and 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

McDonald, S., Buck, J., Song, Q., and Li, Z. (2023). Genome-wide association study reveals novel loci and a candidate gene for resistance to frogeye leaf spot (*Cercospora sojina*) in soybean. *Molecular Genetics and Genomics*, 298, 441-454. https://doi.org/10.1007/s00438-022-01986-z

Meis, S.J., Fehr, W.R., and Schnebly, S.R. (2003). Seed source effect on field emergence of soybean lines with reduced phytate and raffinose saccharides. *Crop Science*, 43(4), 1336–1339. https://doi.org/10.2135/cropsci2003.1336

Messina, M. and Messina, V. (2010). The role of soy in vegetarian diets. *Nutrients*, 2(8), 855-888. https://doi.org/10.3390/nu2080855

Moseley, D., Mozzoni, L., Kaler, A., Mason, R.E., Shi, A., Orazaly, M., et al. (2021). Evaluation of genetic diversity and association mapping for seed weight and size in vegetable soybean germplasm. *Crops Science*, 61(5), 3516-3528. https://doi.org/10.1002/csc2.20588

Nian, H., Wang, J.L., Yang, Q.K., Chen, Y., Lan, X.Y., and Chui, Y.G. (1996). Effect of ecological environments on fat acids contents in seeds of various soybean cultivars. *Soybean Science*, 15(1), 35-41.

Ogawa, S., Valencia, M.O., Lorieux, M., Arbelaez, J.D., McCouch, M.I., and Selvaraj, M.G. (2016). Identification of QTLs associated with agronomic performance under nitrogen-deficient conditions using chromosome segment substitution lines of a wild rice relative, *Oryza rufipogon*. *Acta Physiologiae Plantarum*, 38, 103. https://doi.org/10.1007/s11738-016-2119-5

Patil, G., Valliyodan, B., Deshmukh, R., Prince, S., Nicander, B., Zhao, M., et al. (2015). Soybean (*Glycine max*) SWEET gene family: Insights through comparative genomics, transcriptome profiling and whole genome re-sequence analysis. *BMC Genomics*, 16, 520. https://doi.org/10.1186/s12864-015-1730-y

Patil, G., Vuong, T.D., Kale, S., Valliyodan, B., Deshmukh, R., Zhu, C., et al. (2018). Dissecting genomic hotspots underlying seed protein, oil, and sucrose content in an interspecific mapping population of soybean using high-density linkage mapping. *Plant Biotechnology Journal*, 16, 1939-1953. https://doi.org/10.1111/pbi.12929

Peterbauer, T.P., and Richter, A. (2001). Biochemistry and physiology of raffinose family oligosaccharides and galactosyl cyclitols in seeds. *Seed Science Research*, 11(3), 185-197. https://doi.org/10.1079/SSR200175

Piper, E.L., and Boote, K.J. (1999). Temperature and cultivar effects on soybean seed oil and protein concentrations. *Journal of the American Oil Chemists' Society*, 76(10), 1233-1241. https://doi.org/10.1007/s11746-999-0099-y

Pipolo, A.E., Sinclair, T.R., and 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

Qin, P., Wang, T., and 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

Rani, K., Kumar, M., Razzaq, A., Ajay, B.C., Kona, P., Bera, S.K., et al. (2023). Recent advances in molecular marker technology for QTL mapping in plants. In S.H. Wani, D. Wang, and G.P. Singh (Eds.), *QTL Mapping in Crop Improvement: Present Progress and Future Perspectives* (pp. 1-15). Academic Press.

Rathore, P., Dumenyo, K., and Taheri, A. (2022). Genome-Wide Association study for root system architecture traits in field Soybean [*Glycine max* (L.) Merr.]. *Authorea*. http://doi.org/10.22541/au.167146736.60840320/v1

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

Redekar, N.R., Glover, N.M., Biyashev, R.M., Ha, B.K., Raboy, V., and Maroof, M.A.S. (2020). Genetic interactions regulating seed phytate and oligosaccharides in soybean (*Glycine max* L.). *PLOS ONE*, 15(6), e0235120. https://doi.org/10.1371/journal.pone.0235120

Riaz, A., Raza, Q., Kumar, A., Dean, D., Chiwina, K., Phiri, T.M., et al. (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

Ritchie, H. (2021). Is our appetite for soy driving deforestation in the Amazon? Our World in Data. Available at https://ourworldindata.org/soy#article. Accessed 17 December 2023.

Rizzo, G., and Baroni, L. (2018). Soy, soy foods and their role in vegetarian diets. *Nutrients*, 10(1), 43. https://doi.org/10.3390/nu10010043

Rosset, M., Prudencio, S.H., and 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

Ruan, Y.L. (2014). Sucrose metabolism: Gateway to diverse carbon use and sugar signaling. *Annual Review of Plant Biology*, 65, 33-67. https://doi.org/10.1146/annurev-arplant-050213-040251

Saghai Maroof, M.A., and Buss, G.R. (2008). Low phytic acid, low stachyose, high sucrose soybean lines. U.S. Patent 12/033,830, field Feb.19, 2008.

Salari, M.W., Ongom, P.O., Thapa, R., Nguyen, H.T., Vuong, T.D., and Rainey, K.M. (2021). Mapping QTL controlling soybean seed sucrose and oligosaccharides in a single family of soybean nested association mapping (SoyNAM) population. *Plant Breeding*, 140(1), 110-122. https://doi.org/10.1111/pbr.12883

Saleem, A., Roldán-Ruiz, I., Aper, J., and Muylle, H. (2022). Genetic control of tolerance to drought stress in soybean. *BMC Plant Biology*, 22, 615. https://doi.org/10.1186/s12870-022-03996-w

Saravitz, D.M., Pharr, D.M., and Carter, T.E. (1987). Galactinol synthase activity and soluble sugars in developing seeds of four soybean genotypes. *Plant Physiology*, 83(1), 185-189. https://doi.org/10.1104/pp.83.1.185

Sebastian, S.A., Kerr, P.S., Pearlstein, R.W., and 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.

Seck, W., Torkamaneh, D., and Belzile, F. (2020). Comprehensive Genome-Wide Association Analysis Reveals the Genetic Basis of Root System Architecture in Soybean. *Frontiers in Plant Science*, 11, 590740. https://doi.org/10.3389/fpls.2020.590740

Sedivy, E.J., Wu, F., and Hanzawa, Y. (2017). Soybean domestication: the origin, genetic architecture and molecular bases. *New Phytologist*, 214(2), 539-553. https://doi.org/10.1111/nph.14418

Semba, R.D., Ramsing, R., Rahman, N., Kraemer, K., and Bloem, M.W. (2021). Legumes as a sustainable source of protein in human diets. *Global Food Security*, 28, 100520. https://doi.org/10.1016/j.gfs.2021.100520

Shi, Y.G., Yang, Y., Piekoszewski, W., Zeng, J.H., Guan, H.N., Wang, B., et al. (2020). Influence of four different coagulants on the physicochemical properties, textural characteristics and flavour of tofu. *International Journal of Food Science and Technology*, 55(3), 1218-1229. https://doi.org/10.1111/ijfs.14357

Shurtleff, W., and Aoyagi, A. (2004). History of world soybean production and trade. In History of soybeans and soyfoods, 1100 B.C. to the 1980s. Soyfoods Center, Lafayette, California. Available at: (https://www.soyinfocenter.com/HSS/production_and_trade1.php). Accessed 16 January 2024.

Singer, W.M., Lee, Y., Shea, Z., Vieira, C.C., Lee, D., Li, X., et al. (2023). Soybean genetics, genomics, and breeding for improving nutritional value and reducing antinutritional traits in food and feed. *The Plant Genome*, 16(4), e20415. https://doi.org/10.1002/tpg2.20415

Singh, R.J., and Hymowitz, T. (1999). Soybean genetic resources and crop improvement. *Genome*, 42(4), 605-616. https://doi.org/10.1139/g99-039

Skoneczka, J.A., Saghai Maroof, M.A., Shang, C., and Buss, G.R. (2009). Identification of Candidate Gene Mutation Associated With Low Stachyose Phenotype in Soybean Line PI200508. *Crop Science*, 49, 247-255. https://doi.org/10.2135/cropsci2008.07.0403

Soystats (2023). International: World soybean Production. Available at: http://soystats.com/international-world-soybean-production/. Accessed 23 January 2024.

Stein, O., and Granot, D. (2019). An overview of sucrose synthases in plants. *Frontiers in Plant Science*, 10, 95. https://doi.org/10.3389/fpls.2019.00095

Storz, M.A., Brommer, M., Lombardo, M., and Rizzo, G. (2023). Soy Milk Consumption in the United States of America: An NHANES Data Report. *Nutrients*, 15(11), 2532. https://doi.org/10.3390/nu15112532

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

Sung, M., Van, K., Lee, S., Nelson, R., LaMantia, J., Taliercio, E., et al. (2021). Identification of SNP markers associated with soybean fatty acids contents by genome-wide association analyses. *Molecular Breeding*, 41, 27. https://doi.org/10.1007/s11032-021-01216-1

Teleky, B., Martau, A.G., Ranga, F., Chetan, F., and Vodnar, D.C. (2020). Exploitation of Lactic Acid Bacteria and Baker's Yeast as Single or Multiple Starter Cultures of Wheat Flour Dough Enriched with Soy Flour. *Biomolecules*, 10(5), 778. https://doi.org/10.3390/biom10050778

Thomas, J.M.G., Boote, K.J., Allen, L.H., Gallo-Meagher, M., and Davis, J.M. (2003). Elevated Temperature and Carbon Dioxide Effects on Soybean Seed Composition and Transcript Abundance. *Crop Science*, 43(4), 1548-1557. https://doi.org/10.2135/cropsci2003.1548

TrendEconomy (2023). Soya beans (soybeans) Imports and Exports 2022. Available at https://trendeconomy.com/data/commodity_h2/1201. Accessed 3 January 2024.

USDA (2020). USDA announces $100 million for American biofuels infrastructure (Release No. 0237.20). Available at https://www.usda.gov/media/press-releases/2020/05/04/usda-announces-100-million-american-biofuels-infrastructure. Accessed 1 December 2023.

USDA (2022). World Agricultural Production. Circular Series, WAP 2-22, February 2022. Available at https://apps.fas.usda.gov/PSDOnline/Circulars/2022/02/production.pdf. Accessed 9 January 2024.

Uysal, N., Acik, G., and Tasdelen, M.A. (2017). Soybean oil based thermoset networks via photoinduced CuAAC click chemistry. *Polymer International*, 66(7), 999-1004. https://doi.org/10.1002/pi.5346

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

Vicente-Carbajosa, J., and Carbonero, P. (2005). Seed maturation: developing an intrusive phase to accomplish a quiescent state. *The International Journal of Developmental Biology*, 49, 645-651. https://doi.org/10.1387/ijdb.052046jc

Vieira, C.C., and Chen, P. (2021). The numbers game of soybean breeding in the United States. *Crop Breeding and Applied Biotechnology*, 21(S), e387521S10. https://doi.org/10.1590/1984-70332021v21Sa23

Vieira, C.C., Zhou, J., Usovsky, M., Vuong, T., Howland, A., Lee, D., et al. (2022). Exploring Machine Learning Algorithms to Unveil Genomic Regions Associated With Resistance to Southern Root-Knot Nematode in Soybeans. *Frontiers in Plant Science*, 13, 883280. https://doi.org/10.3389/fpls.2022.883280

Vuong, T.D., Sonah, H., Deshmukh, R., Kadam, S., Meinhardt, C.G., Nelson, R., et al. (2015). Genetic architecture of cyst nematode resistance revealed by genome-wide association study in soybean. *BMC Genomics*, 16, 593-604. http://doi.org/10.1186/s12864-015-1811-y

Wang, J., Chu, S., Zhang, H., Zhu, Y., Cheng, H., and Yu, D. (2016). Development and application of a novel genome-wide SNP array reveals domestication history in soybean. *Scientific Reports*, 6(1), 20728. https://doi.org/10.1038/srep20728

Wang, Y., Chen, P., and 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., Yu, D., Morota, G., Dhakal, K., Singer, W., Lord, N., et al. (2023). Genome-wide association analysis of sucrose concentration in soybean (*Glycine max* L.) seed based on high-throughput sequencing. *The Plant Genome*, 13(3), e20059. https://doi.org/10.1002/tpg2.20059

Wilson, R.F. (2008). Soybean: Market driven research needs. In: G. Stacey (Ed.), *Genetics and Genomics of Soybean* (pp. 3-15). Springer.

Wu, C., Mozzoni, L.A., Moseley, D., Hummer, W., Ye, H., Chen, P., et al. (2019). Genome-wide association mapping of flooding tolerance in soybean. *Molecular Breeding*, 40, 4. https://doi.org/10.1007/s11032-019-1086-0

Xu, D., Abe, J., Gai, J., and Shimamoto, Y. (2002). Diversity of chloroplast DNA SSRs in wild and cultivated soybeans: evidence for multiple origins of cultivated soybean. *Theoretical and Applied Genetics*, 105, 645-653. https://doi.org/10.1007/s00122-002-0972-7

Yang, G., Li, W., Fan, C., Liu, M., Liu, J., Liang, W., et al. (2022b). Genome-wide association study uncovers major genetic loci associated with flowering time in response to active accumulated temperature in wild soybean population. *BMC Genomics*, 23, 749. https://doi.org/10.1186/s12864-022-08970-2

Yang, H., Bian, Y., Huang, L., Lan, Q., Ma, L., Li, X., et al. (2022a). Effects of replacing fish meal with fermented soybean meal on the growth performance, intestinal microbiota, morphology and disease resistance of largemouth bass (*Micropterus salmoides*). *Aquaculture Reports*, 22, 100954. https://doi.org/10.1016/j.aqrep.2021.100954

Yin, L., Zhang, H., Tang, Z., Xu, J., Yin, D., Zhang, Z., et al. (2020). rMVP: a memory-efficient, visualization-enhanced, and parallel-accelerated tool for genome wide association study. *Genomics, Proteomics, and Bioinformatics*, 19(4), 619-628. https://doi.org/10.1101/2020.08.20.258491

Yoosefzadeh-Najafabadi, M., Rajcan, I., and Vazin, M. (2022). High-throughput plant breeding approaches: Moving along with plant-based food demands for pet food industries. *Frontiers in Veterinary Science*, 9, 991844. https://doi.org/10.3389/fvets.2022.991844

Yu, D., Lin, T., Sutton, K., Lord, N., Carneiro, R., Jin, Q., et al. (2021). Chemical Compositions of Edamame Genotypes Grown in Different Locations in the US. *Frontiers in Sustainable Food Systems*, 5, 620426. https://doi.org/10.3389/fsufs.2021.620426

Yuan, W., Wu, Z., Zhang, Y., Yang, R., Wang, H., Kan, G., et al. (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

Zhang, S., Hao, D., Zhang, S., Zhang, D., Wang, H., Du, H., et al. (2021). Genome wide association mapping for protein, oil and water-soluble protein contents in soybean. *Molecular Genetics and Genomics*, 296, 91-102. https://doi.org/10.1007/s00438-020-01704-7

Zeng, A., Chen, P., Shi, A., Wang, D., Zhang, B., Orazaly, M., et al. (2014). Identification of quantitative trait loci for sucrose content in soybean seed. *Crop Science*, 54(2), 554-564. https://doi.org/10.2135/crops ci2013.01.0036

Zhou, Z., Jiang, Y., Wang, Z., Gou, Z., Lyu, J., Li, W., et al. (2015). Resequencing 302 wild and cultivated accessions identifies genes related to domestication and improvement in soybean. *Nature Biotechnology*, 33, 408-414. https://doi.org/10.1038/nbt.3096

CHAPTER 2

Novel genetic resources associated with sucrose and stachyose content through genome-wide association study in soybean [*Glycine max* (L.) Merr.]

Lee, D., Lara, L., Moseley, D., Vuong, T.D., Shannon, G., Xu, D., and Nguyen, H.T. (2023). Novel genetic resources associated with sucrose and stachyose content through genome-wide association study in soybean [*Glycine max* (L.) Merr.]. *Frontiers in Plant Science*, 14, 1294659. https://doi.org/10.3389/fpls.2023.1294659

ABSTRACT

The nutritional value of soybean [*Glycine max* (L.) Merr.] for animals is influenced by soluble carbohydrates, such as sucrose and stachyose. Although sucrose is nutritionally desirable, stachyose is an antinutrient causing diarrhea and flatulence in non-ruminant animals. We conducted a genome-wide association study (GWAS) of 220 soybean accessions using 21,317 single nucleotide polymorphisms (SNPs) from the SoySNP50K iSelect Beadchip data to identify significant SNPs associated with sucrose and stachyose content. Seven significant SNPs were identified for sucrose content across chromosomes (Chrs.) 2, 8, 12, 17, and 20, while thirteen significant SNPs were identified for stachyose content across Chrs. 2, 5, 8, 9, 10, 13, 14, and 15. Among those significant SNPs, three sucrose-related SNPs on Chrs. 8 and 17 were novel, while twelve stachyose-related SNPs on Chrs. 2, 5, 8, 9, 10, 13, 14, and 15 were novel. Based on Phytozome, STRING, and GO annotation, 17 and 24 candidate genes for sucrose and stachyose content, respectively, were highly associated with the carbohydrate metabolic pathway. Among these, the publicly available RNA-seq Atlas database highlighted four candidate genes associated with sucrose (*Glyma.08g361200* and *Glyma.17g258100*) and stachyose (*Glyma.05g025300* and *Glyma.13g077900*) content, which had higher gene expression levels in developing seed and multiple parts of the soybean plant. The results of this study will extend knowledge of the molecular mechanism and genetic basis underlying sucrose and stachyose content in soybean seed. Furthermore, the novel candidate genes and SNPs can be valuable genetic resources that soybean breeders may utilize to modify carbohydrate profiles for animal and human usage.

INTRODUCTION

Soybean [*Glycine max* (L.) Merr.] is one of the most economically valuable crops, providing high protein meal and vegetable oil for human and animal diets worldwide. Annually, almost 76% of global soybean production is used to feed livestock for the meat and dairy industries (Ritchie and Roser, 2021). In comparison, 20% and 4% of production are directly used for the human diet and biofuel industry, respectively (Ritchie and Roser, 2021). In 2021, the U.S., the world's second soybean producer, produced almost 46 million metric tons of soybean meal, which was fed to poultry (61.2%), swine (18.0%), dairy (13.4%), beef (5.1%), and others (2.3%) (American Soybean Association, 2021). Recently, soy-based products have garnered more attention from the global vegan population as a high-value protein substitute for animal meats (Qin et al., 2022). Also, the number of companion pet owners who prefer plant-based pet foods, such as soybean, over animal-based products has been growing due to animal welfare, ethical, and moral concerns (Yoosefzadeh-Najafabadi et al., 2022). In the U.S., the pet industry is an important market that has been booming for decades, of which approximately 66% of the U.S. households own a pet and spend almost 58.1 billion dollars on pet food and treats annually (APPA, 2023).

Soybean seed typically consists of 40% protein, 20% oil, and 15% soluble carbohydrates on a dry weight basis (Hsu et al., 1973). While soybean meal is high in crude protein content with well-balanced amino acids for animal feeds, some antinutrients in soluble carbohydrates significantly reduce feed efficiency for non-ruminant animals, including poultry, swine, dogs, cats, and humans (Jo et al., 2018; Cunicelli et al., 2019; Jo et al., 2019). Sucrose is the only soluble carbohydrate nutritionally beneficial to produce

metabolizable energy. In contrast, raffinose and stachyose make up the raffinose family of oligosaccharides (RFOs), known as antinutrients causing diarrhea and flatulence in non-ruminant animals (Liu, 1997; Guillon and Champ, 2002; Karr-Lilienthal et al., 2005). Non-ruminant animals lack a-galactosidase in their digestive systems, in which undigested RFOs pass through the upper intestine and are fermented by anaerobic microbes in the lower intestine. This produces methane, hydrogen, and carbon dioxide that cause gastric discomfort and a significant loss of energy efficiency from the soybean meal (Coon et al., 1990; Le et al., 2020; Salari et al., 2021). Since the largest soybean meal consumers are non-ruminant animals, developing new soybean cultivars with high sucrose and low RFOs is crucial to improve digestibility and feed efficiency. In addition, increased sucrose content in soybean seeds is also essential to improve the sweet flavor of soy-based products, such as tofu, edamame, and soymilk (Rosset et al., 2012; Sui et al., 2020; Wang et al., 2023).

Compared to other seed compositional traits in soybean, such as oil and protein, a relatively smaller number of quantitative trait loci (QTL) for soluble carbohydrates have been identified and reported through genetic linkage analysis. Historically, Maughan et al. (2000) first reported 17 QTL related to sucrose across chromosomes (Chrs.) 5, 7, 8, 13, 15, 19, and 20 using 149 F_2 individuals from an interspecific cross between *G. max* and *G. soja*. Other studies in South Korea reported four sucrose- and two oligosaccharides-related QTL on Chrs. 2, 11, and 19 and Chrs. 2 and 19, respectively. Two common QTL on Chrs. 2 and 19 were found for both traits in the recombinant inbred line (RIL) population (Kim et al., 2005). A year later, Kim et al. (2006) identified two sucrose- and four oligosaccharides-related QTL on Chrs. 12 and 16 and Chrs. 6, 12, 16, and 19, respectively. Two QTL on Chrs. 12 and 16 were identified for both traits. Skoneczka et al. (2009)

analyzed two F_2 populations and identified a major QTL on Chr. 6 for sucrose and stachyose, which explained 76% and 88% of the phenotypic variations, respectively. Saghai Maroof and Buss (2008) found a major QTL related to both sucrose and stachyose on Chr. 11. Using the F_2 population derived from the same high sucrose soybean line used by Saghai Maroof and Buss (2008), Wang et al. (2014) identified three sucrose-related QTL on Chrs. 7, 11, and 20 and two stachyose-related QTL on 11 and 12. The QTL on Chr. 11 was in the same genetic region reported by Saghai Maroof and Buss (2008). Zeng et al. (2014) reported three novel QTL on Chrs. 5, 9, and 16, explaining 46%, 10%, and 8% of sucrose variation, respectively. Akond et al. (2015) found three sucrose- and four stachyose-associated QTL on Chrs. 3, 9, and 15 and 1, 6, 12, and 14, respectively. Patil et al. (2018) identified three QTL on Chrs. 6, 16, and 20, and a major QTL on Chr. 8 for sucrose using an interspecific population derived from a cross between *G. max* and *G. soja* accessions.

Genome-wide association study (GWAS) is a valid alternative to genetic linkage analysis to understand the genetic basis of quantitative traits by examining a significant association between genetic markers and a trait of interest. To date, GWAS has been successfully applied in soybean research to discover and characterize key traits, such as seed protein and oil (Hwang et al., 2014; Li et al., 2019; Zhang et al., 2021), amino acids (Lee et al., 2019; Yuan et al., 2021), fatty acids (Liu et al., 2020; Sung et al., 2021), disease resistance (Vuong et al., 2015; Ravelombola et al., 2020; Vieira et al., 2022; McDonald et al., 2023), abiotic stress tolerance (Kaler et al., 2017; Wu et al., 2019; Saleem et al., 2022), agronomic traits (Ayalew et al., 2022; Cao et al., 2022; Yang et al., 2022), and root system (Seck et al., 2020; Rathore et al., 2022; Kim et al., 2023). However, only a few studies have

implemented GWAS for soluble carbohydrates in soybean seeds (Lu et al., 2022; Xu et al., 2022).

In this study, a diverse panel of 220 soybean accessions and 21,317 polymorphic single nucleotide polymorphisms (SNPs) were used to conduct GWAS to identify significant marker-trait associations for sucrose and stachyose through a mixed linear model (MLM). Among the three main soluble carbohydrates, raffinose was excluded from this study due to the little phenotypic variation and non-Gaussian distribution in the accession panel. The gene function, protein interaction, biochemical pathway, and gene expression of potential candidate genes associated with sucrose and stachyose content were further studied.

MATERIALS AND METHODS

Accession panel selection and field experimental design

A diverse panel of 220 soybean plant introductions (PIs) was selected based on the 100-seed weight (> 23 g) and relevant maturity groups (MGs) from the USDA-ARS Soybean Germplasm Collection (https://www.ars-grin.gov/) (Supplementary Table 1). The panel included four MGs, III, IV, V, and VI, that originated from six countries (China, Japan, North Korea, South Korea, Taiwan, and the United States). The panel was grown at the Arkansas Agricultural Research and Extension Center (36.06 °N 94.16 °W) in Fayetteville, AR, and the Rice Research and Extension Center (34.47 °N 91.41 °W) in Stuttgart, AR, in 2014 (FAY_14 and STU_14, respectively) and 2015 (FAY_15 and STU_15, respectively). Each accession was planted in single 3-m rows spaced 75 cm apart in a randomized complete block design (RCBD) with two replications. In 2020 and 2021, the accessions were grown at the Fisher Delta Research, Extension, and Education Center (FDREEC) (36.42 °N 89.70 °W) in Portageville, MO (POR_20 and POR_21) and the Bradford Research and Extension Center (BREC) (38.89 °N 92.19 °W) in Columbia, MO (COL_21). Ten seeds of each accession were planted in 75 cm wide rows in hill plots spaced 30 cm apart at the FDREEC and 60 cm apart at the BREC. The experiments at FDREEC and BREC were planted in a RCBD design with two replications.

Soluble carbohydrates phenotyping

Soybean seeds of each accession in each replication were harvested at maturity. Ten seeds were sampled per plot to quantify soluble carbohydrates in the Soybean Genetics & Genomics Laboratory under the supervision of Dr. Henry Nguyen at the University of

Missouri, Columbia. The content of soluble carbohydrates was measured using the established High-Performance Liquid Chromatography (HPLC) protocol described by Valliyodan et al. (2015). Briefly, around 1 g of soybean seeds was ground using Thomas Wiley Mini-Mill (Arthur Thomas Co., Chadds Ford, PA, USA) fitted with a 20-mesh screen. The soybean powder was then lyophilized for 48 hours using a Labconco freeze-dry system (Labconco, Kansas City, MO, USA). Precisely, 90.25 (± 0.15) mg of dried soybean powder was mixed with 900 mL HPLC-grade water in a 2 mL centrifuge tube. Each tube was incubated at 55°C, agitated at 200 rpm for an hour, and then vortexed for 30 seconds. After 20 minutes under room temperature, 900 mL HPLC-grade acetonitrile was added to each tube. Next, the suspension was centrifuged for 30 minutes at a 14.0 × 1000 min^{-1} × g speed. The supernatant was diluted five times with 65% HPLC-grade acetonitrile to prepare the final sample. The final samples were loaded on the Agilent HPLC-ELSD (Evaporative Light Scattering Detection) 120 series (Agilent, Santa Clara, CA, USA), equipped with the Prevail Carbohydrate ES columns (5 mm 250 × 4.6 mm) and guard columns (7.5 × 4.6 mm) (Grace Davison Discovery Sciences, Deerfield, IL, USA). Standard mixtures were prepared in HPLC-grade water with 50, 100, 300, 500, and 1000 mg/ mL concentrations to create calibration curves.

Statistical analysis

Analysis of variance (ANOVA) was conducted using the GLM procedure of SAS software version 9.4 with 'Genotype' within 'Maturity group' as fixed effects and 'Environment', 'Maturity group', 'Genotype × Environment', and 'Replication' as random effects. The best linear unbiased prediction (BLUP) was computed using the *lmer* function

in R software to minimize the effects of environmental variation and used as an additional environment in GWAS analysis. Pearson's correlation coefficients between sucrose and stachyose were calculated using the *chart.Correlation* function in R software across environments and within each environment. The significant difference in mean values across seven environments between favorable and unfavorable alleles was determined using the PROC ANOVA function in the SAS software.

Genotype data processing

The SoySNP50K iSelect Beadchip data for the 220 PI lines were obtained from Soybase (https://www.soybase.org/). The SNPs with a minor allele frequency (MAF) less than 0.05 were removed using GAPIT (Lipka et al., 2012). A total of 21,317 SNPs were used for GWAS in this study. The number of filtered SNPs mapped across 20 soybean chromosomes ranged from 693 on Chr. 20 to 1,677 on Chr. 18, with an average of 1,066 SNPs (Supplementary Figure 1).

Population structure analysis and linkage disequilibrium estimation

Population structure was analyzed using the STRUCTURE software version 2.3.4 (Pritchard et al., 2000). The hypothetical number of subpopulations (K) from 2 to 9 was set with five independent iterations. For each run, the burn-in iteration and Markov Chain Monte Carlo replication were set at 10,000 and 25,000, respectively. Principal component analysis (PCA) was performed using TASSEL software version 5.0 (Bradbury et al., 2007). The Linkage disequilibrium (LD) block was calculated by computing correlation coefficients (r^2) for all pairwise marker comparisons and visualized using Haploview

software to identify potential candidate genes (Barrett et al., 2005). The kinship matrix was also generated by centered-IBS methods using TASSEL software.

Genome-wide association study

Principal components (PCs) and a kinship matrix were incorporated in MLM as covariates to correct for population structure and cryptic relatedness. The MG was additionally used as a categorical covariate since the MG effects were significant on sucrose and stachyose based on ANOVA. The significant threshold of SNP-trait association at a *P*-value = 1.0×10^{-3} was suggested to identify consistent and significant SNPs across all environments. The GWAS was conducted for each environment and BLUP. Significant SNPs were determined when the SNP was identified in three or more environments. If multiple significant SNPs were detected within the same LD block, the most consistent SNP with the highest $-\log_{10}(P)$ value was selected.

The gene function, protein interaction, and biochemical pathway of candidate genes

All potential candidate genes within the LD block of each significant SNP were obtained using *Glycine max* cv. Williams 82 reference-genome gene models version 2.0 in Soybase. Also, the predicted amino acid sequences of the potential candidate genes were obtained from Soybase and used to study the protein interaction network using the STRING database (https://string-db.org/) (Szklarczyk et al., 2019). Relevant candidate genes were selected based on the metabolic studies of carbohydrates. The most adjacent candidate gene was selected if no relevant candidate gene was found within the LD block. Gene ontology (GO) annotation was obtained to confirm the biological processes, cellular components,

and molecular functions of the relevant candidate genes using Database for Annotation, Visualization and Integrated Discovery (DAVID) bioinformatics resources (http://david.ncifcrf.gov/) (Sherman et al., 2021).

Tissue-specific gene expression analysis

The RNA-Seq Atlas data, publicly available on Soybase (http://soybase.org/soyseq/), were used to compare the gene expression levels between the relevant candidate genes in 14 different soybean plant tissues, including young leaf, flower, 1-cm pod, pod shell 10 days after flowering (DAF), pod shell 14 DAF, seed 10 DAF, seed 14 DAF, seed 21 DAF, seed 25 DAF, seed 28 DAF, seed 35 DAF, seed 42 DAF, root, and nodule (Severin et al., 2010). The raw gene expression counts were normalized using a Reads Per Kilobase of transcript per Million mapped reads (RPKM) method. Highly expressed candidate genes were determined based on at least two gene expression levels of more than 10 RPKM in developing seed tissues (https://www.ebi.ac.uk/). Another large set of RNA-Seq databases of gene-level transcript abundances (https://soyatlas.venanciogroup.uenf.br/) was used for additional exploration in the differential gene expression across 19 parts of the soybean plant, including cotyledon, embryo, endosperm, epicotyl, flower, hypocotyl, leaf, nodule, petiole, pod, radicle, root, seed, seed coat, seedling, shoot, suspensor, unknown, and whole plant (Almeida-Silva et al., 2023). The raw gene expression counts were normalized using a Transcripts Per Million (TPM) method. Highly expressed candidate genes were determined based on the total gene expression level of more than 1000 TPM across all parts (https://www.ebi.ac.uk/).

RESULTS

Evaluations of phenotypic data

The panel of 220 soybean accessions was tested for sucrose and stachyose content in seven environments. A summary of the phenotypic values of each environment is shown in Table 1. The highest mean value for sucrose was 7.3% in POR_20, with a range of 4.5% - 9.5%, while the lowest was 5.8% in STU_15, with a range of 4.1% - 8.4%. The coefficients of variation (CV) ranged from 11.0% (POR_20) to 14.4% (STU_14), and Shapiro-Wilk (w) values ranged from 0.974 (STU_15) to 0.995 (FAY_15). The highest mean value for stachyose was 4.5% in STU_14, with a range of 2.2% - 6.3%, while the lowest mean was 3.5% in FAY_15, with a range of 1.0% - 5.6%. The CV ranged from 16.2% (POR_21) to 25.5% (FAY_15), and Shapiro-Wilk (w) values ranged from 0.940 (STU_15) to 0.991 (FAY_15). Pearson's correlations among environments for sucrose content varied from 0.26 (between STU_14 and POR_21) to 0.77 (between FAY_14 and STU_14) (Supplementary Figure 2). Pearson's correlations among environments for stachyose content varied from 0.64 (between STU_14 and POR_21 and between STU_15 and COL_21) to 0.93 (between FAY_15 and POR_20). Also, Pearson's correlation between sucrose and stachyose content was estimated to be significantly negative in three environments (FAY_14, FAY_15, and STU_15) and significantly positive in one environment (STU_14) (Supplementary Table 2). The ANOVA showed significant effects for 'Genotype' within maturity groups, 'Environment', 'Maturity group', and 'Genotype × Environment' for sucrose and stachyose content but no significant effect for 'Replication' (Supplementary Table 3).

Population structure

The cryptic relatedness among 220 soybean accessions, estimated by kinship matrix and visualized in the heat map, indicated two distinct subpopulations (1 and 2) (Figure 1A). The PCA plot demonstrated that the two components accounted for 29.1% (PC1) and 8.4% (PC2) of genetic variation, which differentiate the two subpopulations (Figure 1B). The first subpopulation had 116 accessions with origins from China (4), Japan (79), North Korea (1), South Korea (22), Taiwan (6), and the United States within MG III (30), IV (43), V (21), and VI (22). The second subpopulation had 104 accessions with origins from China (2), Japan (2), North Korea (5), and South Korea (95) within MG III (1), IV (100), V (2), and VI (1). STRUCTURE analysis also suggested that the optimal number of subpopulations (K) was two among all genotypes (Figure 1C).

Association study for sucrose and stachyose content

Manhattan and quantile-quantile (Q-Q) plots for each environment for sucrose and stachyose content are shown in Figures 2 and 3, respectively. Across all environments, a total of 88 and 89 SNPs were associated with sucrose and stachyose content, respectively (Supplementary Table 4).

Seven out of 88 SNPs, identified in three or more environments for sucrose content, were considered significant SNPs. The seven significant SNPs were located on Chrs. 2 (ss715581183), 8 (ss715602502), 12 (ss715613179), 17 (ss715627820 and ss715627853), and 20 (ss715636857 and ss715637428) (Figure 2; Table 2). The SNP ss715602502 located at 47,286,262 bp on Chr. 8, identified in five environments (FAY_14, FAY_15, STU_15, COL_21, and BLUP), was the most consistent SNP for sucrose content. The SNPs

ss715613179 located at 5,486,355 bp on Chr. 12 were identified in four environments: FAY_15, COL_21, POR_21, and BLUP. The SNP ss715627853 located at 41,440,620 bp on Chr. 17 were also identified in four environments: FAY_14, COL_21, POR_21, and BLUP (Table 2; Supplementary Table 4). The highest -$\log_{10}$ (*P*) value (5.2) was found on the SNP ss715627853 on Chr. 17. The MAF of the significant SNPs ranged from 0.05 (ss715627820) to 0.39 (ss715636857) (Table 2).

Thirteen out of 89 SNPs, identified in three or more environments for stachyose content, were considered significant SNPs. The thirteen significant SNPs were located on Chrs. 2 (ss715583079 and ss715583119), 5 (ss715592340, ss715592442, and ss715591198), 8 (ss715601133), 9 (ss715603880 and ss715639178), 10 (ss715606330), 13 (ss715614101 and ss715615716), 14 (ss715617675), and 15 (ss715622806) (Figure 3; Table 2). The SNP ss715592442 located at 2,369,980 bp on Chr. 5, identified in seven environments (FAY_14, STU_14, FAY_15, STU_15, POR_20, COL_21, and BLUP), was the most consistent SNP for stachyose content. The SNP ss715583119 located at 44,448,179 bp on Chr. 2 for stachyose content was identified in six environments: FAY_14, FAY_15, STU_15, POR_20, COL_21, and BLUP (Table 2; Supplementary Table 4). The highest -$\log_{10}$ (*P*) value (5.0) was found on the SNP ss715592442 on Chr. 5. The MAF of the significant SNPs ranged from 0.08 (ss715606330) to 0.47 (ss715639178) (Table 2).

Allelic effect of significant SNPs for sucrose and stachyose content

The allelic effects of significant SNPs for sucrose and stachyose content were tested using mean values of favorable and unfavorable alleles across seven environments (Figure 4). An allele conferring higher sucrose content was designated a favorable allele. In contrast,

an allele conferring lower stachyose content was designated a favorable allele. The favorable alleles of five SNPs, ss715581183 (Chr. 2), ss715613179 (Chr. 12), ss715627820 (Chr. 17), ss715627853 (Chr. 17), and ss715637428 (Chr. 20), were related to significantly higher sucrose content than unfavorable alleles (*P*-value < 0.001) (Figure 4A). However, two SNPs, ss715602502 (Chr. 8) and ss715636857 (Chr. 20), showed no significant difference between favorable and unfavorable alleles for sucrose content. The favorable alleles of eight SNPs, ss715583079 (Chr. 2), ss715583119 (Chr. 2), ss715601133 (Chr. 8), ss715603880 (Chr. 9), ss715606330 (Chr. 10), ss715614101 (Chr. 13), ss715617675 (Chr. 14), and ss715622806 (Chr. 15), were related to significantly lower stachyose content than unfavorable alleles (*P*-value < 0.001) (Figure 4B). However, the five SNPs, ss715592340 (Chr. 5), ss715592442 (Chr. 5), ss715591198 (Chr. 5), ss715639178 (Chr. 9), and ss715615716 (Chr. 13), showed no significant difference between favorable and unfavorable alleles for stachyose content. SNPs having no significant allelic effects, however, still showed favorable phenotypic trends across environments (Supplementary Figure 3). To evaluate the pyramiding effects of favorable alleles of significant SNPs, the variations of sucrose and stachyose content across seven environments were compared in the different numbers of favorable alleles (Table 3). The results showed that sucrose contents ranged from 5.7% ± 0.7 (none) to 7.9% ± 0.4 (six favorable alleles), and stachyose contents ranged from 2.0% ± 0.2 (13 favorable alleles) to 5.3% ± 0.0 (one favorable allele).

Prediction of potential candidate genes for sucrose and stachyose content

A total of 107 and 155 genes located within the LD blocks of significant SNPs were considered potential candidate genes for sucrose and stachyose, respectively

(Supplementary Table 5). The estimated size of LD blocks varied from 6 to 528 Kbp with an average of 186 Kbp, which is slightly longer than that of previously tested cultivated soybeans (~150 Kbp) (Lam et al., 2010) (Supplementary Table 5).

Out of 107 genes, 17 relevant candidate genes for sucrose content were closely associated with carbohydrate metabolism (Table 4). Since no candidate gene related to carbohydrate metabolism was located within the LD block of ss715637428 for sucrose content, the most adjacent candidate gene to the significant SNP, *Glyma.20g099600*, was selected. The largest number of relevant candidate genes (seven) were found in the LD block of ss715602502 on Chr. 8 related to three sugar transporter-related genes, *Glyma.08g360400*, *Glyma.08g360500*, and *Glyma.08g361200*, and four other genes, *Glyma.08g356800*, *Glyma.08g357200*, *Glyma.08g358700*, and *Glyma.08g358800*, that interact with carbohydrate metabolism-related proteins (polygalacturonase, 1D- myo-inositol 3-kinase, mannosyl-oligosaccharide α-1,3-glucosidase, and UDP-sugar pyrophosphorylase, respectively). Also, four genes, *Glyma.12g072800*, *Glyma.20g017400*, *Glyma.20g018000*, and *Glyma.20g018200*, are involved in carbohydrate metabolism, while five genes, *Glyma.02g129200*, *Glyma.17g257800*, *Glyma.17g258100*, *Glyma.17g260300*, and *Glyma.17g260400*, have functional interactions with carbohydrate metabolism-related proteins.

Furthermore, GO annotation of the 17 relevant candidate genes for sucrose content confirmed six biological processes, including GO:0045490 (pectin catabolic process), GO:0006012 (galactose metabolic process), GO:0008643 (carbohydrate transport), GO:0006096 (glycolytic process), GO:0006004 (fucose metabolic process), and GO:0005975 (carbohydrate metabolic process) (Supplementary Table 6). Seven molecular

functions were also confirmed, including GO:0030570 (pectate lyase activity), GO:0004335 (galactokinase activity), GO:0051119 (sugar transmembrane transporter activity), GO:0005366 (myo-inositol:proton symporter activity), GO:0004340 (glucokinase activity), GO:0005536 (glucose binding), and GO:0042973 (glucan endo-1,3-beta-D-glucosidase activity) (Supplementary Table 6).

Out of 155 genes, 24 relevant candidate genes for stachyose content were closely associated with carbohydrate metabolism (Table 4). Since a stachyose-associated significant SNP, ss715639178, showed no LD block range, the most adjacent gene, *Glyma.09g124300*, was selected as a candidate gene. Also, due to no candidate gene related to carbohydrate metabolism within the LD blocks of ss715603880 and ss715615716, the most adjacent genes, *Glyma.09g044100* and *Glyma.13g077900*, were selected, respectively. The largest number of relevant candidate genes (four) were found in the LD block of ss715614101 on Chr. 13, where carbohydrate metabolism-related genes, *Glyma.13g128300* (sugar kinase), *Glyma.13g128400* (sugar kinase), *Glyma.13g132700* (D-myo-inositol (1,4,5)-trisphosphate degradation), and *Glyma.13g133800* (UDP-glucuronate decarboxylase), are located. Also, nine genes, *Glyma.02g255100*, *Glyma.05g024800*, *Glyma.05g025300*, *Glyma.05g025400*, *Glyma.05g026800*, *Glyma.08g028600*, *Glyma.10g040700*, *Glyma.14g113100*, and *Glyma.15g275400*, are involved in carbohydrate metabolism, while eight genes, *Glyma.02g255400*, *Glyma.02g258000*, *Glyma.02g258200*, *Glyma.05g027100*, *Glyma.05g166800*, *Glyma.10g040000*, *Glyma.15g275300*, and *Glyma.15g276700*, have functional interactions with carbohydrate metabolism-related proteins.

Furthermore, GO annotation of the 24 relevant candidate genes for stachyose content confirmed seven biological processes, including GO:0006486 (protein glycosylation), GO:0009969 (xyloglucan biosynthetic process), GO:0019323 (pentose catabolic process), GO:0006012 (galactose metabolic process), GO:0019252 (starch biosynthetic process), GO:0042732 (D-xylose metabolic process), and GO:0005975 (carbohydrate metabolic process) (Supplementary Table 6). Seven molecular functions were also confirmed, including GO:0016757 (glycosyltransferase activity), GO:0008107 (galactoside 2-alpha-L-fucosyltransferase activity), GO:0004335 (galactokinase activity), GO:0019200 (carbohydrate kinase activity), GO:0048040 (UDP-glucuronate decarboxylase activity), GO:0030247 (polysaccharide binding), and GO:0004650 (polygalacturonase activity) (Supplementary Table 6).

Tissue-specific gene expression analysis for relevant candidate genes

Two genes, *Glyma.08g361200* and *Glyma.17g258100*, related to sucrose content had distinctively higher gene expression levels of 160 and 273 RPKM, respectively, than other candidate genes in the developing seeds from Seed 10 DAF to Seed 42 DAF (Figure 5A). Furthermore, these two genes showed the highest total TPM values (1,041 and 1,061 TPM, respectively) across 19 plant parts in soybean (Figure 5B). *Glyma.17g258100* especially had considerably higher gene expression levels in seed and seed coats, while the gene expression of *Glyma.08g361200* was distributed throughout the plant. The gene expression data of *Glyma.12g072800* and *Glyma.17g260300* (Figure 5A) and *Glyma.08g358800* and *Glyma.12g072800* (Figure 5B) was not available.

Four genes, *Glyma.05g025300*, *Glyma.09g044100*, *Glyma.13g077900*, and *Glyma.13g133800*, related to stachyose content, showed relatively higher gene expression levels (58, 76, 99, 71 RPKM, respectively) than other candidate genes in the developing seeds from Seed 10 DAF to Seed 42 DAF (Figure 5C). *Glyma.05g025300* had the highest total gene expression level (1,625 TPM), followed by *Glyma.13g077900* (1,450 TPM) across 19 plant parts in soybean (Figure 5D). *Glyma.05g025300* had higher gene expression levels in leaf and shoots, while *Glyma.13g077900* had a higher gene expression level in the epicotyl (embryonic shoot). The gene expression data of *Glyma.13g128400* and *Glyma.14g113100* (Figure 5C) and *Glyma.05g026800*, *Glyma.09g124300*, *Glyma.13g128400*, and *Glyma.14g113100* (Figure 5D) was not available.

DISCUSSION

In the current study, the sucrose and stachyose content showed significant phenotypic variations across the environments studied (Table 1). Similar to other seed composition traits, such as protein and oil, soluble carbohydrates are also influenced by diverse environmental factors, especially temperature during pod-filling stages (Bandillo et al., 2015; Bilyeu and Wiebold, 2016). In a previous study, Bilyeu and Wiebold (2016) reported cooler temperatures during pod-filling stages increased sucrose content while decreasing stachyose content in soybean. Among the environments studied, the POR_20 showed relatively cooler temperatures compared to the others during the pod-filling stages, while the STU_15 showed relatively warmer temperatures than the others (Supplementary Figure 4). The sucrose content of the POR_20 was the highest, and that of the STU_15 was the lowest, which followed the well-known relationship between temperature and sucrose content (Jo et al., 2018; Jo et al., 2019). On the other hand, the trend of stachyose content across environments was less influenced by temperature conditions, which the result was in good agreement with earlier reports by Kumar et al. (2010); Jo et al. (2018), who concluded that stachyose content was more genotype dependent. In future investigations, more geographically dispersed locations will be required to extensively test potential environmental influences on soluble carbohydrates.

Assessing multiple environment phenotypic datasets, a subset of 24 soybean accessions with either/both high sucrose and/or low stachyose content was selected for further genetic analysis and breeding applications (Table 5). Accessions PI 536547 B, PI 561288, and PI 549065 exhibited desirable soluble carbohydrate profiles with high sucrose (> 7.7%) and low stachyose content (< 2.3%), which are greatly beneficial to human and

animal consumption (Jo et al., 2018; Sui et al., 2020). Field evaluations conducted in 2014 and 2015 showed these PI lines had larger seed sizes (> 29.0 g per 100 seeds) and high protein content (> 42.0%) on a dry weight basis. In particular, larger seed sizes and high sucrose and high protein content are essential quality parameters in soy-food production, including edamame, miso, and tofu (Konovsky et al., 1994; Zeipina et al., 2017; Jegadeesan and Yu, 2020). Interestingly, the soybean accessions with higher sucrose in this panel mainly originated from Japan. This indicated the historical selection in Japan was conducted based on the taste of soybean products for human uses, such as edamame, miso, natto, soy sauce, and tofu, of which sucrose is a main contributor to the sweetness of soybean products (Kaga et al., 2012; Rosset et al., 2012; Sui et al., 2020; Wang et al., 2023).

The current bottleneck of narrowed genetic diversity has been addressed by soybean breeders for decades. Breeding and commercializing soybean cultivars with desirable soluble carbohydrate profiles has been slowed due to the limitation of genetic sources, although the importance has been continuously expressed to meet global demand. In addition to those PI lines mentioned above, we employed two additional accessions, PI 506937 and PI 506593, with desirable carbohydrate profiles for developing bi-parental mapping populations in an on-going effort to characterize the genetic architecture of carbohydrate composition traits. We anticipated these PI lines could be useful as novel genetic resources for breeders to develop new specialty soybean cultivars with modified soluble carbohydrate content.

For decades, different analysis models have been developed and utilized for GWAS. Among these, MLM is one of the most popular models, which incorporates population structure and kinship as covariates to reduce false positives and increase the statistical

power in identifying significant marker-trait associations (Yu et al., 2005; Ficht et al., 2022). In the current study, the MG effect was significant on sucrose and stachyose content (Supplementary Table 3). Therefore, the corresponding MG of each genotype was additionally incorporated as a categorical covariate to reduce the potential bias derived from the MG differences (Supplementary Table 1). The relatively smaller panel size (< 300) could be a limitation for GWAS, although it contains four MGs and six origins. To declare a significant SNP, we tested a typical Bonferroni Correction using a stringent threshold ($-\log_{10}(P) = 5.6$); however, no significant SNP was detected. Thus, a general consensus value of 0.001 ($-\log_{10}(P) = 3.0$) was used as a significant cut-off value to detect significant SNPs (Hwang et al., 2014).

Using phenotypic data of each environment and estimated BLUP, seven significant SNPs for sucrose content were identified across five chromosomes (Chrs. 2, 8, 12, 17, and 20) (Figure 2 and Table 2). Among them, four SNPs were confirmed to be close to previously reported sucrose-related genomic regions. Hu et al. (2023) reported a significant SNP ($-\log_{10}(P) = 3.3$) associated with fructose, which was located in the same region as ss715581183 at 13,523,639 bp on Chr. 2. Fructose is one of the monosaccharides that comprise the sucrose molecule. Lu et al. (2022) reported a significant SNP ($-\log_{10}(P) = 8.8$) associated with total soluble carbohydrate at 5,036,567 bp on Chr. 12, which was mapped approximately 449 Kbp downstream of ss715613179 (5,486,355 bp). Total soluble carbohydrates were positively correlated with sucrose in soybean (Hou et al., 2009). Also, Kim et al. (2006) reported a sucrose-related QTL (‘Seed Sucrose 3-4’ at Soybase, LOD = 8.4) associated with the marker Satt442 at 6,390,806 - 6,391,062 bp on Chr. 12, which is mapped approximately 904 Kbp upstream of ss715613179. Based on only 110 markers

used by Kim et al. (2006), the distance of less than a million bp between the two markers was considered significant. Patil et al. (2018) reported a sucrose-related QTL, qSUC_20 (LOD = 3.8), at 2,386,021 - 2,558,940 bp on Chr. 20, which was mapped approximately 478 Kbp upstream of ss715636857 (1,907,881 bp). Wang et al. (2023) also reported a sucrose-related SNP ($-\log_{10}(P)$ = 2.9) at 34,981,501 bp on Chr. 20, which was mapped approximately 690 Kbp upstream of ss715637428 (34,286,637 bp). Another marker, Satt270 (35,362,576 - 35,362,794 bp), on Chr. 20 was reported as an associated marker to a sucrose-related QTL (LOD = 1.9). It was mapped approximately 1 Mbp upstream of ss715637428 (Wang et al., 2014). Three SNPs, ss715602502 on Chr. 8, ss715627820 and ss715627853 on Chr. 17, are novel for sucrose content in soybean.

Thirteen significant SNPs for stachyose content were identified across eight chromosomes (Chrs. 2, 5, 8, 9, 10, 13, 14, and 15) (Figure 3; Table 2). Unlike sucrose, only one SNP was confirmed to be close to the previously reported stachyose/oligosaccharides-related genomic regions. Kim et al. (2005) reported a stachyose-related QTL marker ('Seed Oligosaccharide 1-1' at Soybase, LOD = 7.7) associated with the marker Satt546 at 40,699,300 - 40,699,539 bp on Chr. 2, which was mapped approximately 3.5 Mbp downstream of ss715583079 (44,214,908 bp). The other 12 SNPs identified in this study have not been previously reported and were considered novel SNPs for stachyose content in soybean. Historically, soybean breeding programs have been more interested in sucrose than stachyose, resulting in more than twice as many sucrose-related QTL/SNPs published in the Soybase (https://soybase.org).

In the current study, ss715602502 on Chr. 8 was the most consistent SNP significantly associated with sucrose content, followed by ss715613179 on Chr. 12 and

ss715627853 on Chr. 17 (Figure 2; Supplementary Table 4). However, ss715602502 had no significant allelic effect on sucrose content (Figure 4A). This phenomenon has been explained by the fact that not all variants are causally associated with a trait, and the association can be indirect (MacArthur et al., 2014; Gallagher and Chen-Plotkin, 2018; Schaid et al., 2019). Despite no significant allelic effect of ss715602502, the estimated LD block contained seven candidate genes underlying carbohydrate transport/metabolism (Table 4; Supplementary Tables 5 and 6). Among these, two genes, *Glyma.08g360400* and *Glyma.08g360500*, were reported to belong to the SWEET (Sugars Will Eventually be Exported Transporters) gene family (SWEET25 and SWEET26, respectively), which play essential roles in transporting glucose molecules across a membrane (Patil et al., 2015). The most consistent significant SNP for stachyose content, ss715592442 on Chr. 5, was identified in seven environments. The SNP ss715583119 on Chr. 2 was identified in six environments (Figure 3; Supplementary Table 4). Like sucrose, ss715592442 and ss715583119 had no significant allelic effect on stachyose content. The estimated LD blocks contained carbohydrate metabolism-related genes, including glycosyltransferase family 64 protein c4 (*Glyma.05g026800*), sucrose synthase (*Glyma.05g027100*), glycosyltransferases (*Glyma.02g258000*), and clathrin propeller repeat (*Glyma.02g258200*) (Figure 4B; Table 4). The biological process of *Glyma.02g258200* is to route acidic a-galactosidases to protein storage vacuoles, and it is known to facilitate the accumulation of RFOs during seed development in pea (Blöchl et al., 2008). Therefore, integrative post-GWAS analyses, including protein interaction, metabolic pathway, biological function, cellular component, and molecular function, are crucial to comprehensively understand the variants/candidate genes identified in GWAS (Jia et al., 2011).

Two candidate genes associated with sucrose, *Glyma.08g361200* and *Glyma.17g258100*, showed high gene expression levels in the developing seed tissues and total expression levels across 19 different parts of the soybean plant (Figures 5A, B). The molecular function of *Glyma.08g361200* is myo-inositol:proton symporter activity, which is responsible for transferring myo-inositol from one side of a membrane to the other (Supplementary Table 6). Myo-inositol is essential for galactinol biosynthesis, and galactinol plays a key role in the chain elongation of sucrose to RFOs (Saravitz et al., 1987; Irvine and Schell, 2001). However, no information was available for *Glyma.17g258100* in the GO database. Instead, based on Phytozome and STRING database, it is responsible for gibberellin-regulated protein and protein interaction with xyloglucan endotransglucosylase/hydrolase (Table 4). Xyloglucan endotransglucosylase/hydrolase catalyzes the cleavage of the main chain of xyloglucan molecules until only glucose, xylose, and galactose are produced, of which glucose is an essential component of the sucrose metabolism (Buckeridge, 2023).

Two candidate genes, *Glyma.05g025300* and *Glyma.13g077900*, showed high total gene expression levels for stachyose content in developing seed tissues and total expression levels across 19 different parts of the soybean plant (Figures 5C and 5D). Unlike sucrose, these genes showed relatively higher gene expression levels in leaf, shoot, and epicotyl (embryonic shoot) than in seed and seed coats. This suggested that stachyose content may be affected by genes expressed not only in the developing seed but also in aerial parts, including leaf and shoot. The raffinose synthase 3 gene (*Glyma.05g003900*, Wm82.a2.v1), widely used in soybean breeding programs to reduce RFOs, also has the highest expression level in shoots, followed by petioles and leaves in soybean (Supplementary Figure 5). The

gene, *Glyma.05g025300*, is responsible for the carbohydrate metabolic process (GO:0005975), while *Glyma.13g077900* is responsible for the proton-transporting V-type ATPase (GO:0033179) (Supplementary Table 6). In plants, V-type ATPase plays an important role in regulating signaling events in response to diverse environmental stimuli (Elmore and Coaker, 2011). Therefore, this confirmed that stachyose is also known to play an important role in protection against abiotic stress, such as salinity, cold, and drought in plants (Mundree et al., 2002; ElSayed et al., 2014; Zhang et al., 2019; Yan et al., 2022).

To date, three major genes, *Glyma.05g003900* (raffinose synthase 3, RS3), *Glyma.06g179200* (raffinose synthase 2, RS2), and *Glyma.11g238800* (D-myo-inositol-3-phosphate synthase 1, MIPS1), primarily responsible for increased sucrose and reduced RFOs, have been identified. Even though molecular markers related to these genes have been utilized for marker-assisted selection (MAS) breeding, a limited number of soybean cultivars for human and animal consumption have been developed and commercialized. Recently, Benson Hill, Inc. (https://bensonhill.com/) announced the release of a new soybean cultivar with ultra-high protein and low RFOs content, coupled with other improved agronomic traits. In this study, we reported new genomic regions harboring candidate genes associated with increased sucrose and reduced stachyose, which did not overlap with the known genes, RS2, RS3, and MIPS1. Therefore, the promising soybean germplasms and novel genetic sources newly identified in this study would be critical to broadening germplasm resources and enriching our understanding of the genetic architecture for the development of new soybean cultivars for animal feed efficiency and natural sweetness in human consumption. Future studies, including genome-editing

technologies, a fine-mapping strategy, and molecular marker development, will be required for the functional validation of the novel findings in this study.

CONCLUSION

This study identified three and 12 novel SNPs associated with sucrose and stachyose content, respectively, through a GWAS using 220 soybean accessions. Four novel candidate genes for sucrose (*Glyma.08g361200* and *Glyma.17g258100*) and stachyose (*Glyma.05g025300* and *Glyma.13g077900*) content were further identified. We also reported three promising lines (PI 536547 B, PI 561288, and PI 549065) as germplasm resources that can be valuable for developing new soybean cultivars with desirable soluble carbohydrate profiles. The novel discoveries in this study will extend the current knowledge of the genetic basis underlying sucrose and stachyose content in soybean seed. Overall, new genetic resources will benefit soybean breeders in modifying carbohydrate profiles to meet the high demand for animal and human consumption.

REFERENCES

Akond, M., Liu, S., Kantartzi, S.K., Meksem, K., Bellaloui, N., Lightfoot, D.A., et al. (2015). Quantitative trait loci underlying seed sugars content in "MD96-5722" by "Spencer" Recombinant inbred line population of soybean. *Food and Nutrition Sciences*, 6, 964–973. https://doi.org/10.4236/fns.2015.611100

Almeida-Silva, F., Pedrosa-Silva, F., and Venancio, T.M. (2023). The soybean expression atlas v2: a comprehensive database of over 5000 RNA-seq samples. *bioRxiv*, 28, 538661. https://doi.org/10.1101/2023.04.28.538661

American Soybean Association (2021). Soystats. Available at: http://soystats.com. Accessed 12 May 2023.

APPA American Pet Products Association (2023). *Pet Industry Market Size, Trends & Ownership Statistics.* Stamford, CT: APPA. Available at: https://www.americanpetproducts.org/press_industrytrends.asp. Accessed 12 May 2023.

Ayalew, H., Schapaugh, W., Vuong, T., and Nguyen, H.T. (2022). Genome-wide association analysis identified consistent QTL for seed yield in a soybean diversity panel tested across multiple environments. *The Plant Genome*, 15(4), e20268. https://doi.org/10.1002/tpg2.20268

Bandillo, N., Jarquin, D., Song, Q., Nelson, R., Cregan, P., Specht, J., et al. (2015). A population structure and genome-wide association analysis on the USDA soybean germplasm collection. *The Plant Genome*, 8, 1–13. https://doi.org/10.3835/plantgenome2015.04.0024

Barrett, J.C., Fry, B., Maller, J., and Daly, M.J. (2005). Haploview: analysis and visualization of LD and haplotype maps. *Bioinformatics*, 21(2), 263–265. https://doi.org/10.1093/bioinformatics/bth457

Bilyeu, K., and 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

Blöchl, A., Peterbauer, T., Hofmann, J., and Richter, A. (2008). Enzymatic breakdown of raffinose oligosaccharides in pea seeds. *Planta*, 228, 99-110. https://doi.org/10.1007/s00425-008-0722-4

Bradbury, P.J., Zhang, Z., Kroon, D.E., Casstevens, T.M., Ramdoss, Y., and Buckler, E.S. (2007). TASSEL: software for association mapping of complex traits in diverse samples. *Bioinformatics*, 23, 2633-2635. https://doi.org/10.1093/bioinformatics/btm308

Buckeridge, M.S. (2023). The diversity of plant carbohydrate hydrolysis in nature and technology. In R. Goldbeck and P. Poletto (Eds.), *Polysaccharide-Degrading Biocatalysts* (pp. 55-74). Academic Press. https://doi.org/10.1016/B978-0-323-99986-1.00015-6

Cao, Y., Jia, S., Chen, L., Zeng, S., Zhao, T., and Karikari, B. (2022). Identification of major genomic regions for soybean seed weight by genome-wide association study. *Molecular Breeding*, 42, 38. https://doi.org/10.1007/s11032-022-01310-y

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

Cunicelli, M.J., Bhandari, H.S., Chen, P., Sams, C.E., Rouf Mian, M.A., Mozzoni, L.A., et al. (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

Elmore, J.M., and Coaker, G. (2011). The Role of the Plasma Membrane H+-ATPase in Plant–Microbe Interactions. *Molecular Plant*, 4(3), 416-427. https://doi.org/10.1093/mp/ssq083

ElSayed, A.I., Rafudeen, M.S., and Golldack, D. (2014). Physiological aspects of raffinose family oligosaccharides in plants: protection against abiotic stress. *Plant Biology*, 16(1), 1-8. https://doi.org/10.1111/plb.12053

Ficht, A., Bruce, R., Torkamaneh, D., Grainger, C.M., Eskandari, M., and 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

Gallagher, M.D., and Chen-Plotkin, A.S. (2018). The Post-GWAS Era: From Association to Function. *The American Journal of Human Genetics*, 102, 717-730. https://doi.org/10.1016/j.ajhg.2018.04.002

Guillon, F., and Champ, M.M.J. (2002). Carbohydrate fractions of legumes: Uses in human nutrition and potential for health. *British Journal of Nutrition*, 88, S293-306. https://doi.org/10.1079/BJN2002720

Hou, A., Chen, P., Alloatti, J., Li, D., Mozzoni, L., Zhang, B., et al. (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

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

Hu, L., Wang, X., Zhang, J., Florez-Palacios, L., Song, Q., and Jiang, G.L. (2023). Genome-Wide Detection of Quantitative Trait Loci and Prediction of Candidate Genes for Seed Sugar Composition in Early Mature Soybean. *International Journal of Molecular Science*, 24(4), 3167. https://doi.org/10.3390/ijms24043167

Hwang, E.Y., Song, Q., Jia, G., Specht, J.E., Hyten, D.L., Costa, J., et al. (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

Irvine, R.F. and Schell, M.J. (2001). Back in the water: the return of the inositol phosphates. *Nature Review Molecular Cell Biology*, 2, 327-338. https://doi.org/10.1038/35073015

Jegadeesan, S., and Yu, K. (2020). Food grade soybean breeding, current status and future directions. In M. Hasanuzzaman (Ed.), *Legume Crops: Prospects, Production and Uses* (pp. 51-78). IntechOpen.

Jia, P., Zheng, S., Long, J., Zheng, W., and Zhao, Z. (2011). dmGWAS: dense module searching for genome-wide association studies in protein–protein interaction networks. *Bioinformatics*, 27(1), 95-102. https://doi.org/10.1093/bioinformatics/btq615

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

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

Kaga, A., Shimizu, T., Watanabe, S., Tsubokura, Y., Katayose, Y., Harada, K., et al. (2012). Evaluation of soybean germplasm conserved in NIAS genebank and development of mini core collections. *Breeding Science*, 61(5), 566-592. https://doi.org/10.1270/jsbbs.61.566

Kaler, A.S., Ray, J.D., Schapaugh, W.T., King, C.A., and Purcell, L.C. (2017). Genome-wide association mapping of canopy wilting in diverse soybean genotypes. *Theoretical and Applied Genetics*, 130, 2203-2217. https://doi.org/10.1007/s00122-017-2951-z

Karr-Lilienthal, L.K., Grieshop, C.M., Spears, J.K., and Fahey, Jr. G.C. (2005). Amino acid, carbohydrate, and fat composition of soybean meals prepared at 55 commercial U.S. soybean processing plants. *Journal of Agricultural and Food Chemistry*, 53, 2146-2150. https://doi.org/10.1021/jf048385i

Kim, H.K., Kang, S.T., Cho, J.H., Choung, M.G., and 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., and 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, S.H., Tayade, R., Kang, B.H., Hahn, B.S., Ha, B.K., and Kim, Y.H. (2023). Genome-Wide Association Studies of Seven Root Traits in Soybean (*Glycine max* L.) Landraces. *International Journal of Molecular Sciences*, 24(1), 873. https://doi.org/10.3390/ijms24010873

Konovsky, J., Lumpkin, T.A., and McClary, D. (1994). EDAMAME: THE VEGETABLE SOYBEAN. In A.D. O'Rourke (Ed.), *Understanding the Japanese food and agrimarket: a multifaceted opportunity* (pp.173-181). Haworth Press.

Kumar, V., Rani, A., Goyal, L., Dixit, A.K., Manjaya, J.G., Dev, J., et al. (2010). Sucrose and raffinose family oligosaccharides (RFOs) in soybean seeds as influenced by genotype and growing location. *Journal of Agricultural and Food Chemistry*, 58, 5081-5085. http://doi.org/10.1021/jf 903141s

Lam, H.M., Xu, X., Liu, X., Chen, W., Yang, G., Wong, F.L., et al. (2010). Resequencing of 31 wild and cultivated soybean genomes identifies patterns of genetic diversity and selection. *Nature Genetics*, 42, 1053-1059. https://doi.org/10.1038/ng.715

Le, H., Nguyen, N.H., Ta, D.T., Le T.N.T., Bui, T.P., Le, N.T., et al. (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, S., Van, K., Sung, M., Nelson, R., LaMantia, J., McHale, L.K., et al. (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, 1639-1659. https://doi.org/10.1007/s00122-019-03304-5

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

Lipka, A.E., Tian, F., Wang, Q., Peiffer, J., Li, M., Bradbury, P.J., et al. (2012). GAPIT: Genome association and prediction integrated tool. *Bioinformatics*, 28, 2397-2399. https://doi.org/10.1093/bioinformatics/bts444

Liu, K. (1997). Chemistry and nutritional value of soybean components. In K. Liu (Ed.), *Soybeans* (pp. 25-113). Springer US. https://doi.org/10.1007/978-1-4615-1763-4_2

Liu, X., Qin, D., Piersanti, A., Zhang, Q., Miceil, C., and Wang, P. (2020). Genome-wide association study identifies candidate genes related to oleic acid content in soybean seeds. *BMC Plant Biology*, 20, 399. https://doi.org/10.1186/s12870-020-02607-w

Lu, W., Sui, M., Zhao, X., Jia, H., Han, D., Yan, X., et al. (2022). Genome-Wide Identification of Candidate Genes Underlying Soluble Sugar Content in Vegetable Soybean (*Glycine max* L.) via Association and Expression Analysis. *Frontiers in Plant Science*, 13, 930639. https://doi.org/10.3389/fpls.2022.930639

MacArthur, D.G., Manolio, T.A., Dimmock, D.P., Rehm, H.L., Shendure, J., Abecasis, G.R., et al. (2014). Guidelines for investigating causality of sequence variants in human disease. *Nature*, 508, 469-476. https://doi.org/10.1038/nature13127

Maughan, P.J., Saghai Maroof, M.A., and 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

McDonald, S., Buck, J., Song, Q., and Li, Z. (2023). Genome-wide association study reveals novel loci and a candidate gene for resistance to frogeye leaf spot (*Cercospora sojina*) in soybean. *Molecular Genetics and Genomics*, 298, 441-454. https://doi.org/10.1007/s00438-022-01986-z

Mundree, S.G., Baker, B., Mowla, S., Peters, S., Marais, S., Willigen, C.V., et al. (2002). Physiological and molecular insights into drought tolerance. *African Journal of Biotechnology*, 1(2), 28-38. https://doi.org/10.5897/AJB2002.000-006

Patil, G., Valliyodan, B., Deshmukh, R., Prince, S., Nicander, B., Zhao, M., et al. (2015). Soybean (*Glycine max*) SWEET gene family: insights through comparative genomics, transcriptome profiling and whole genome re-sequence analysis. *BMC Genomics*, 16, 520. http://doi.org/10.1186/s12864-015-1730-y

Patil, G., Vuong, T.D., Kale, S., Valliyodan, B., Deshmukh, R., Zhu, C., et al. (2018). Dissecting genomic hotspots underlying seed protein, oil, and sucrose content in an interspecific mapping population of soybean using high-density linkage mapping. *Plant Biotechnology Journal*, 16, 1939-1953. https://doi.org/10.1111/pbi.12929

Pritchard, J.K., Stephens, M., and Donnelly, P. (2000). Inference of population structure using multilocus genotype data. *Genetics*, 155, 945–959. https://doi.org/10.1093/genetics/155.2.945

Qin, P., Wang, T., and 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

Rathore, P., Dumenyo, K., and Taheri, A. (2022). Genome-Wide Association study for root system architecture traits in field Soybean [*Glycine max* (L.) Merr.]. *Authorea*. http://doi.org/10.22541/au.167146736.60840320/v1

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

Ritchie, H., and Roser, M. (2021). Forests and deforestation (OurworldInData.org). Available at: https://ourworldindata.org/forests-and-deforestation.

Rosset, M., Prudencio, S.H., and 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

Saghai Maroof, M.A., and Buss, G.R. (2008). Low phytic acid, low stachyose, high sucrose soybean lines. U.S. Patent 12/033,830, field Feb.19, 2008.

Salari, M.W., Ongom, P.O., Thapa, R., Nguyen, H.T., Vuong, T.D., and Rainey, K.M. (2021). Mapping QTL controlling soybean seed sucrose and oligosaccharides in a single family of soybean nested association mapping (SoyNAM) population. *Plant Breeding*, 140(1), 110-122. https://doi.org/10.1111/pbr.12883

Saleem, A., Roldán-Ruiz, I., Aper, J., and Muylle, H. (2022). Genetic control of tolerance to drought stress in soybean. *BMC Plant Biology*, 22, 615. https://doi.org/10.1186/s12870-022-03996-w

Saravitz, D.M., Pharr, D.M., and Carter, T.E. (1987). Galactinol synthase activity and soluble sugars in developing seeds of four soybean genotypes. *Plant Physiology*, 83(1), 185-189. https://doi.org/10.1104/pp.83.1.185

Schaid, D.J., Chen, W., and Larson, N.S. (2019). From genome-wide associations to candidate causal variants by statistical fine-mapping. *Nature Reviews Genetics*, 19(8), 491-504. http://doi.org/10.1038/s41576-018-0016-z

Seck, W., Torkamaneh, D., and Belzile, F. (2020). Comprehensive Genome-Wide Association Analysis Reveals the Genetic Basis of Root System Architecture in Soybean. *Frontiers in Plant Science*, 11, 590740. https://doi.org/10.3389/fpls.2020.590740

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

Sherman, B.T., Hao, M., Qiu, J., Jiao, X., Baseler, M.W., Lane, H.C., et al. (2021). DAVID: a web server for functional enrichment analysis and functional annotation of gene lists. *Nucleic Acids Research*, 50(W1), W216-W221. https://doi.org/10.1093/nar/gkac194

Skoneczka, J.A., Saghai Maroof, M.A., Shang, C., and Buss, G.R. (2009). Identification of Candidate Gene Mutation Associated With Low Stachyose Phenotype in Soybean Line PI200508. *Crop Science*, 49, 247-255. https://doi.org/10.2135/cropsci2008.07.0403

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

Sung, M., Van, K., Lee, S., Nelson, R., LaMantia, J., Taliercio, E., et al. (2021). Identification of SNP markers associated with soybean fatty acids contents by genome-wide association analyses. *Molecular Breeding*, 41, 27. https://doi.org/10.1007/s11032-021-01216-1

Szklarczyk, D., Gable, A.L., Lyon, D., Junge, A., Wyder, S., Huerta-Cepas, J., et al. (2019). STRING v11: protein–protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets. *Nucleic Acids Research*, 47, D607-D613. https://doi.org/10.1093/nar/gky1131

Valliyodan, B., Shi, H., and Nguyen, H.T. (2015). A Simple Analytical Method for High-Throughput Screening of Major Sugars from Soybean by Normal-Phase HPLC with Evaporative Light Scattering Detection. *Chromatography Research International*, 2015, 8. https://doi.org/10.1155/2015/757649

Vieira, C.C., Zhou, J., Usovsky, M., Vuong, T., Howland, A.D., Lee, D., et al. (2022). Exploring Machine Learning Algorithms to Unveil Genomic Regions Associated With Resistance to Southern Root-Knot Nematode in Soybeans. *Frontiers in Plant Science*, 13, 883280. https://doi.org/10.3389/fpls.2022.883280

Vuong, T.D., Sonah, H., Deshmukh, R., Kadam, S., Meinhardt, C.G., Nelson, R., et al. (2015). Genetic architecture of cyst nematode resistance revealed by genome-wide association study in soybean. *BMC Genomics*, 16, 593-604. http://doi.org/10.1186/s12864-015-1811-y

Wang, Y., Chen, P., and 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., Yu, D., Morota, G., Dhakal, K., Singer, W., Lord, N., et al. (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

Wu, C., Mozzoni, L.A., Moseley, D., Hummer, W., Ye, H., Chen, P., et al. (2019). Genome-wide association mapping of flooding tolerance in soybean. *Molecular Breeding*, 40, 4. https://doi.org/10.1007/s11032-019-1086-0

Xu, W., Liu, H., Li, S., Zhang, W., Wang, Q., Zhang, H., et al. (2022). GWAS and Identification of Candidate Genes Associated with Seed Soluble Sugar Content in Vegetable Soybean. *Agronomy*, 12(6), 1470. https://doi.org/10.3390/agronomy12061470

Yan, S., Liu, Q., Li, W., Yan, J., and Fernie, A.R. (2022). Raffinose Family Oligosaccharides: Crucial Regulators of Plant Development and Stress Responses. *Critical Reviews in Plant Sciences*, 286-303, https://doi.org/10.1080/07352689.2022.2111756

Yang, G., Li, W., Fan, C., Liu, M., Liu, J., Liang, W., et al. (2022). Genome-wide association study uncovers major genetic loci associated with flowering time in response to active accumulated temperature in wild soybean population. *BMC Genomics*, 23, 749. https://doi.org/10.1186/s12864-022-08970-2

Yoosefzadeh-Najafabadi, M., Rajcan, I., and Vazin, M. (2022). High-throughput plant breeding approaches: Moving along with plant-based food demands for pet food industries. *Frontiers in Veterinary Science*, 9, 991844. https://doi.org/10.3389/fvets.2022.991844

Yu, J., Pressoir, G., Briggs, W.H., Bi, I.V., Yamasaki, M., Doebley, J.F., et al. (2005). A unified mixed-model method for association mapping that accounts for multiple levels of relatedness. *Nature Genetics*, 38, 203-208. https://doi.org/10.1038/ng1702

Yuan, W., Wu, Z., Zhang, Y., Yang, R., Wang, H., Kan, G., et al. (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

Zeipina, S., Alsina, I., and Lepse, L. (2017). Insight in edamame yield and quality parameters: A review. *Research for rural development*, 2, 40-45. https://doi.org/10.22616/rrd.23.2017.047

Zeng, A., Chen, P., Shi, A., Wang, D., Zhang, B., Orazaly, M., et al. (2014). Identification of quantitative trait loci for sucrose content in soybean seed. *Crop Science*, 54(2), 554-564. https://doi.org/10.2135/crops ci2013.01.0036

Zhang, S., Hao, D., Zhang, S., Zhang, D., Wang, H., Du, H., et al. (2021). Genome-wide association mapping for protein, oil and water-soluble protein contents in soybean. *Molecular Genetics and Genomics*, 296, 91-102. https://doi.org/10.1007/s00438-020-01704-7

Zhang, Y., Li, D., Zhou, R., Wang, X., Dossa, K., Wang, L., et al. (2019). Transcriptome and metabolome analyses of two contrasting sesame genotypes reveal the crucial biological pathways involved in rapid adaptive response to salt stress. *BMC Plant Biology*, 19, 66. https://doi.org/10.1186/s12870-019-1665-6

TABLES AND FIGURES

Table 1. Summary of sucrose and stachyose content in seven environments.

Trait	Env[1]	Range (%)	Mean (SE)[2]	CV[3]	Shapiro-Wilk (w)	Skewness	Kurtosis
Sucrose	FAY_14	4.5 - 9.1	6.9 (0.057)b	12.2	0.992	-0.01	3.21
	STU_14	3.9 - 8.9	6.4 (0.062)d	14.4	0.990	0.09	2.84
	FAY_15	4.5 - 8.8	6.7 (0.053)c	11.7	0.995	-0.18	2.97
	STU_15	4.1 - 8.4	5.8 (0.049)f	12.4	0.974	0.56	4.03
	POR_20	4.5 - 9.5	7.3 (0.054)a	11.0	0.989	-0.28	3.45
	COL_21	4.1 - 9.0	6.1 (0.056)e	13.5	0.987	0.32	3.18
	POR_21	4.1 - 9.3	6.4 (0.062)d	14.2	0.991	-0.08	2.79
Stachyose	FAY_14	1.1 - 5.6	3.9 (0.061)b	23.2	0.948	-0.85	3.69
	STU_14	2.2 - 6.3	4.5 (0.052)a	16.9	0.984	-0.24	3.46
	FAY_15	1.0 - 5.6	3.5 (0.059)c	25.5	0.991	-0.13	2.88
	STU_15	1.3 - 5.5	3.9 (0.053)b	20.2	0.940	-0.92	4.19
	POR_20	1.4 - 5.4	3.6 (0.051)c	20.9	0.987	-0.27	3.19
	COL_21	2.0 - 5.6	4.1 (0.045)b	16.3	0.984	-0.34	3.21
	POR_21	2.0 - 5.5	4.0 (0.044)b	16.2	0.958	-0.55	3.53

[1]Environments. FAY_14 (Fayetteville in 2014), STU_14 (Stuttgart in 2014), FAY_15 (Fayetteville in 2015), STU_15 (Stuttgart in 2015), POR_20 (Portageville in 2020), COL_21 (Columbia in 2021), POR_21 (Portageville in 2021). [2]Different alphabet letters within the trait indicate mean values are significantly different at $p < 0.05$; SE indicates a standard error. [3]Coefficients of variation.

Table 2. Significant SNPs associated with soybean sucrose and stachyose content identified in three or more environments using a mixed linear model.

Trait	Significant SNP[1]	Chr[2]	Position (bp)	# Env[3]	Allele[4]		-log$_{10}$ (P)	MAF[5]
					Favorable	Unfavorable		
Sucrose	ss715581183	2	13,523,639	3	G	T	3.6 - 4.0	0.37
	ss715602502	8	47,286,262	5	G	T	3.1 - 4.8	0.18
	ss715613179	12	5,486,355	4	C	T	3.2 - 4.0	0.33
	ss715627820	17	41,098,767	3	A	G	3.4 - 4.2	0.05
	ss715627853	17	41,440,620	4	T	C	3.5 - 5.2	0.13
	ss715636857	20	1,907,881	3	G	A	3.1 - 3.7	0.39
	ss715637428	20	34,286,637	3	T	C	3.3 - 3.6	0.07
Stachyose	ss715583079	2	44,214,908	3	G	A	3.3 - 3.7	0.12
	ss715583119	2	44,448,179	6	T	G	3.1 - 3.9	0.11
	ss715592340	5	2,207,089	3	G	A	3.0 - 3.4	0.20
	ss715592442	5	2,369,980	7	C	A	3.2 - 5.0	0.18
	ss715591198	5	35,773,064	4	T	G	3.0 - 3.4	0.16
	ss715601133	8	2,382,910	4	T	C	3.1 - 3.4	0.21
	ss715603880	9	3,771,212	3	C	T	3.0 - 4.0	0.32
	ss715639178	9	30,134,957	4	A	G	3.0 - 3.8	0.47
	ss715606330	10	3,541,231	3	C	T	3.0 - 3.6	0.08
	ss715615716	13	18,379,941	5	C	T	3.1 - 3.9	0.09
	ss715614101	13	24,090,619	3	G	A	3.0 - 3.3	0.14
	ss715617675	14	13,187,218	3	C	T	3.3 - 3.9	0.29
	ss715622806	15	51,630,810	3	G	A	3.0 - 3.2	0.31

[1]Single-nucleotide polymorphism. [2]Chromosome. [3]The number of environments where significant SNP was identified, a total number of environments were eight. [4]The allele conferring higher sucrose content was designated a favorable allele, while the allele conferring lower stachyose content was designated a favorable allele. [5]Minor allele frequency.

Table 3. The phenotypic variations of sucrose and stachyose content across seven environments in the different numbers of favorable alleles of significant SNPs.

Trait	No. of favorable alleles	No. of genotypes	Mean ± SD[1] (%)
Sucrose	0	10	5.7 ± 0.7
	1	50	6.5 ± 0.5
	2	72	6.3 ± 0.6
	3	45	6.5 ± 0.4
	4	32	7.0 ± 0.6
	5	8	7.7 ± 0.5
	6	3	7.9 ± 0.4
Stachyose	1	1	5.3 ± 0.0
	2	3	4.4 ± 0.8
	3	26	4.1 ± 0.5
	4	49	4.1 ± 0.4
	5	82	4.1 ± 0.6
	6	13	3.9 ± 0.4
	7	13	3.4 ± 0.8
	8	10	3.1 ± 0.5
	9	13	3.5 ± 0.8
	10	2	3.0 ± 0.3
	11	3	3.1 ± 1.3
	12	2	2.2 ± 0.2
	13	3	2.0 ± 0.2

[1]Standard deviation.

Table 4. The list of candidate genes associated with carbohydrate metabolic pathways, identified within the linkage disequilibrium block of significant SNPs.

Trait	Significant SNP[1]	Chr[2]	Position (bp)	Candidate gene[3]	Function annotation[4]	Protein interaction[5]
Sucrose	ss715581183	2	13,523,639	*Glyma.02g129200*	Predicted hydrolases of HD superfamily	Xyloglucan galactosyltransferase
	ss715602502	8	47,286,262	*Glyma.08g356800*	Pectin lyase-like superfamily protein	Polygalacturonase
				Glyma.08g357200	Serine-Threonine protein kinase	1D-myo-inositol 3-kinase
				Glyma.08g358700	WD40 repeat-containing protein	Mannosyl-oligosaccharide alpha-1,3-glucosidase
				Glyma.08g358800	D-galactose detoxification	UDP-sugar pyrophosphorylase
				Glyma.08g360400	Sugar efflux transporter for intercellular exchange	12-oxophytodienoic acid reductase
				Glyma.08g360500	Sugar efflux transporter for intercellular exchange	12-oxophytodienoic acid reductase
				Glyma.08g361200	Sugar (and other) transporter	Sucrose transport protein SUC3 isoform -[6]
	ss715613179	12	5,486,355	*Glyma.12g072800*	Glycolysis I and II	
	ss715627820	17	41,098,767	*Glyma.17g257800*	Hexokinase	Glucose-6-phosphate isomerase
				Glyma.17g258100	Gibberellin regulated protein	Xyloglucan endotransglucosylase/hydrolase
	ss715627853	17	41,440,620	*Glyma.17g260300*	Peroxidase	Glycosyltransferases
				Glyma.17g260400	Rare lipoprotein A (RlpA)-like double-psi beta-barrel	Polygalacturonase precursor
	ss715636857	20	1,907,881	*Glyma.20g017400*	GDP-fucose protein O-fucosyltransferase	O-fucosyltransferase 23 isoform
				Glyma.20g018000	GDP-glucose biosynthesis	UTP--glucose-1-phosphate uridylyltransferase isoform X2
				Glyma.20g018200	Glycosyl hydrolases family 17	Glycosyltransferases
	ss715637428	20	34,286,637	*Glyma.20g099600*	Methionyl-tRNA synthetase	Leucine--tRNA ligase
Stachyose	ss715583079	2	44,214,908	*Glyma.02g255100*	Glycosyl transferase	Xyloglucan galactosyltransferase
				Glyma.02g255400	Hs1pro-1 N-terminus	O-fucosyltransferase 20
	ss715583119	2	44,448,179	*Glyma.02g258000*	Response to freezing	Glycosyltransferases
				Glyma.02g258200	Clathrin propeller repeat	Phosphatidylinositol-3,4,5-trisphosphate 3-phosphatase
	ss715592340	5	2,207,089	*Glyma.05g024800*	Galactosyl transferase	Xyloglucan galactosyltransferase
				Glyma.05g025300	Ribulose-phosphate 3 epimerase family	Probable ribose-5-phosphate isomerase
				Glyma.05g025400	Xyloglucan fucosyltransferase	Xyloglucan galactosyltransferase

ss715592442	5	2,369,980	*Glyma.05g026800*	Xylogalacturonan β-1,3-xylosyltransferase	Glycosyltransferase family 64 protein c4
			Glyma.05g027100	alpha/beta-Hydrolases superfamily protein	Sucrose synthase
ss715591198	5	35,773,064	*Glyma.05g166800*	Mitochondrial outer membrane protein 25	Fructose-1,6-bisphosphatase
ss715601133	8	2,382,910	*Glyma.08g028600*	UDP-α-D-galacturonate biosynthesis II (from D-galacturonate)	UDPglucose--hexose-1-phosphate uridylyltransferase / UDP-sugar pyrophosphorylase 1
ss715603880	9	3,771,212	*Glyma.09g044100*	Uncharacterized conserved protein	Condensin-2 complex subunit
ss715639178	9	30,134,957	*Glyma.09g124300*	Photosynthesis	-
ss715606330	10	3,541,231	*Glyma.10g040000*	Glutathione S-transferase	Alpha-1,3-glucosyltransferase
			Glyma.10g040700	Xyloglucan biosynthesis	Xyloglucan 6-xylosyltransferase 2
ss715615716	13	18,379,941	*Glyma.13g077900*	Vacuolar H+-ATPase V0 sector	V-type h+-transporting ATPase subunit d
ss715614101	13	24,090,619	*Glyma.13g128300*	Sugar Kinase	DNA repair protein xrcc4
			Glyma.13g128400	Sugar Kinase	DNA repair protein xrcc4
			Glyma.13g132700	D-myo-inositol (1,4,5)-trisphosphate degradation	Inositol polyphosphate 6-/3-/5-kinase
			Glyma.13g133800	UDP-glucuronate decarboxylase	UDP-sugar pyrophosphorylase 1
ss715617675	14	13,187,218	*Glyma.14g113100*	Polysaccharide binding	-
ss715622806	15	51,630,810	*Glyma.15g275300*	Triose-phosphate Transporter family	Glucose-1-phosphate adenylyltransferase
			Glyma.15g275400	Glycosyl hydrolases family 28	Pectinesterase
			Glyma.15g276700	Uncharacterized protein family	Glucan endo-1,3-beta-glucosidase a6

Note) The most adjacent candidate gene was selected if there was no candidate gene associated with the carbohydrate metabolic pathway.

[1]Single-nucleotide polymorphism. [2]Chromosome. [3]Gene model: *Glycine max* cv. Williams 82 reference-genome gene models version 2.0. [4]Phytozome database. [5]STRING database. [6]Not available.

Table 5. The list of promising soybean germplasm resources to modify soluble carbohydrate profiles for human and animal uses.

Name	Maturity Group	Origin	Sucrose[1]	Stachyose[1]	Protein[2]	Seed size[2]
			Mean ± SD[3] (%)			(g/100 seeds)
PI 506530	VI	Japan	8.8 ± 0.6	4.1 ± 0.5	40.6 ± 1.7	26.0 ± 3.9
PI 506903	IV	Japan	8.6 ± 0.4	3.9 ± 0.4	40.9 ± 1.9	32.6 ± 1.7
PI 506937	IV	Japan	8.2 ± 0.5	3.5 ± 0.7	36.7 ± 1.6	28.7 ± 2.5
PI 229343	IV	Japan	8.0 ± 0.6	4.0 ± 0.5	39.2 ± 1.7	31.5 ± 2.1
PI 507449	IV	Japan	7.9 ± 0.6	2.6 ± 0.8	46.0 ± 1.8	26.8 ± 2.2
PI 507438	VI	Japan	7.8 ± 0.7	3.8 ± 0.8	40.1 ± 2.1	28.7 ± 2.6
PI 398925	VI	South Korea	7.8 ± 0.7	5.1 ± 0.4	42.1 ± 1.8	23.1 ± 1.9
PI 536547 B	III	Taiwan	7.8 ± 0.5	1.9 ± 0.5	43.4 ± 1.4	29.7 ± 3.1
PI 506593	VI	Japan	7.7 ± 0.9	4.8 ± 0.5	41.1 ± 2.1	24.9 ± 3.5
PI 561288	IV	Taiwan	7.7 ± 0.8	2.3 ± 0.7	42.7 ± 1.0	30.3 ± 3.9
PI 548486	VI	Japan	7.7 ± 0.7	5.0 ± 0.6	42.2 ± 2.5	28.2 ± 2.5
PI 416892	III	Japan	7.7 ± 0.6	3.8 ± 0.6	44.9 ± 1.4	32.8 ± 2.9
PI 507523	III	Japan	7.7 ± 0.6	2.5 ± 0.5	46.3 ± 1.0	31.4 ± 2.1
PI 549065	IV	Japan	7.7 ± 0.5	2.1 ± 0.5	43.0 ± 1.2	33.7 ± 3.3
PI 424590 A	IV	South Korea	7.6 ± 0.5	3.8 ± 0.7	38.9 ± 1.2	27.4 ± 1.3
PI 561292 B	IV	Taiwan	7.2 ± 0.5	2.2 ± 1.0	42.7 ± 1.0	33.0 ± 3.2
PI 507273	III	Japan	6.9 ± 0.6	2.5 ± 0.7	40.7 ± 1.2	23.1 ± 1.6
PI 506801 B	III	Japan	6.8 ± 0.7	2.3 ± 0.7	43.2 ± 1.5	28.0 ± 0.9
PI 507487	III	Japan	6.8 ± 0.6	2.4 ± 0.6	43.4 ± 1.5	28.7 ± 2.1
PI 196162	III	Japan	6.8 ± 0.3	2.5 ± 0.5	42.7 ± 0.6	25.2 ± 1.2
PI 506800 A	III	Japan	6.7 ± 0.6	1.8 ± 0.4	45.3 ± 1.1	29.8 ± 0.8
PI 506800 B	III	Japan	6.6 ± 0.5	2.1 ± 0.6	43.8 ± 1.2	29.2 ± 1.4
PI 506799	III	Japan	6.4 ± 0.7	2.2 ± 0.7	44.1 ± 1.4	28.1 ± 1.2
PI 506801 A	III	Japan	5.5 ± 0.5	2.1 ± 0.5	45.4 ± 1.2	26.2 ± 0.9

[1]Traits evaluated across all locations (FAY_14, STU_14, FAY_15, STU_15, POR_20, COL_21, and POR_21). [2]Traits evaluated only across limited locations (FAY_14, STU_14, FAY_15, and STU_15). [3]Standard deviation.

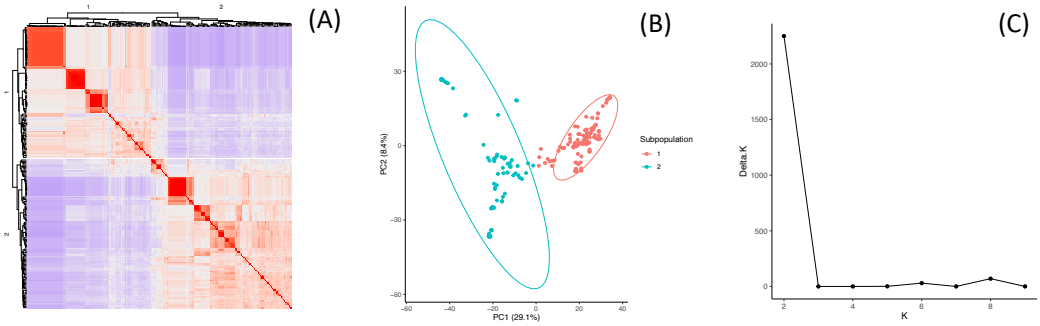


Figure 1. Kinship and population structure analysis of 220 soybean accessions. (A) Kinship matrix visualized in the heat map. (B) PCA plot of the 220 soybean accessions. (C) The optimal number of subpopulations (K) in the accession panel.

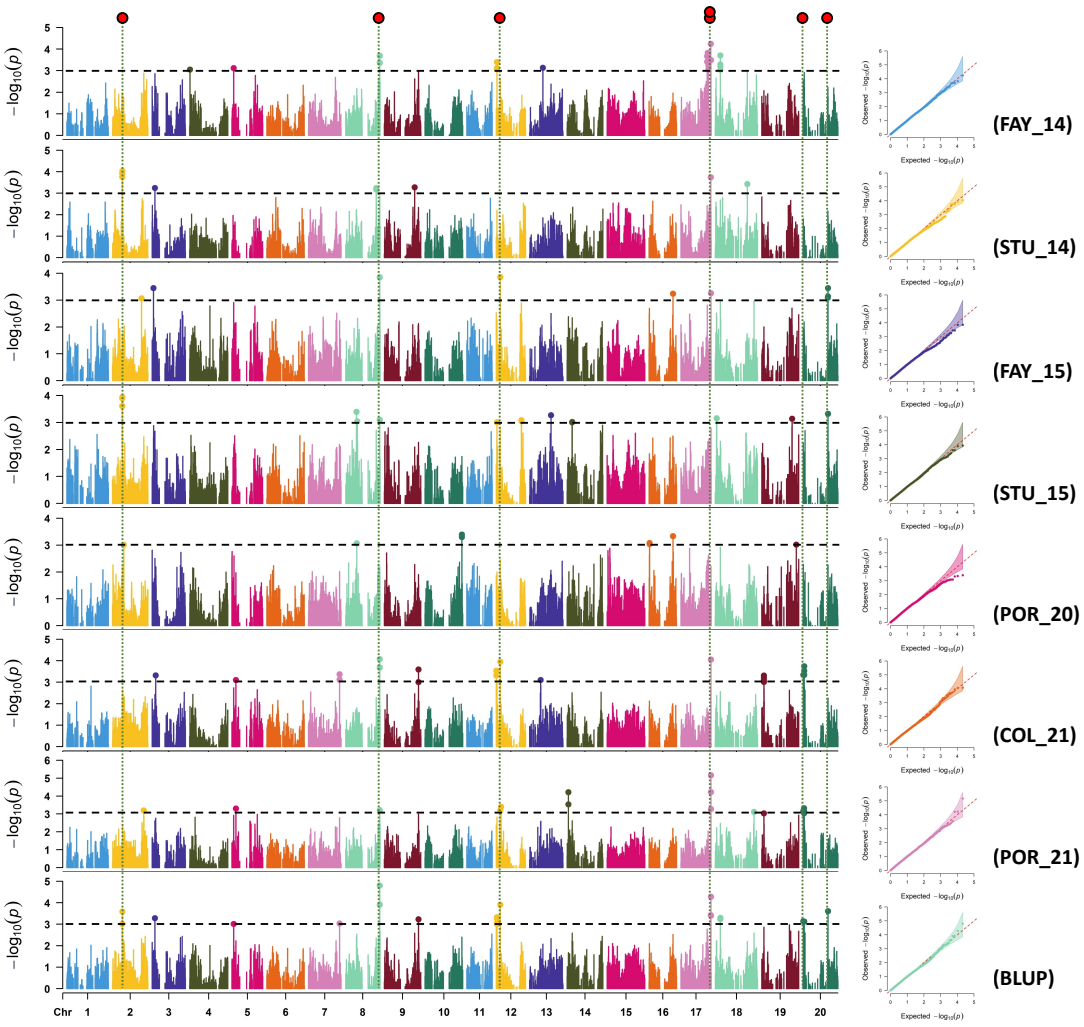


Figure 2. Manhattan and quantile-quantile plots of sucrose for FAY_14 (Fayetteville in 2014), STU_14 (Stuttgart in 2014), FAY_15 (Fayetteville in 2015), STU_15 (Stuttgart in 2015), POR_20 (Portageville in 2020), COL_21 (Columbia in 2021), POR_21 (Portageville in 2021), and BLUP (across seven environments, an additional environment). SNP markers detected in three or more environments were considered significant SNPs using GWAS threshold $-\log_{10}(P) > 3$ (red dots on the top).

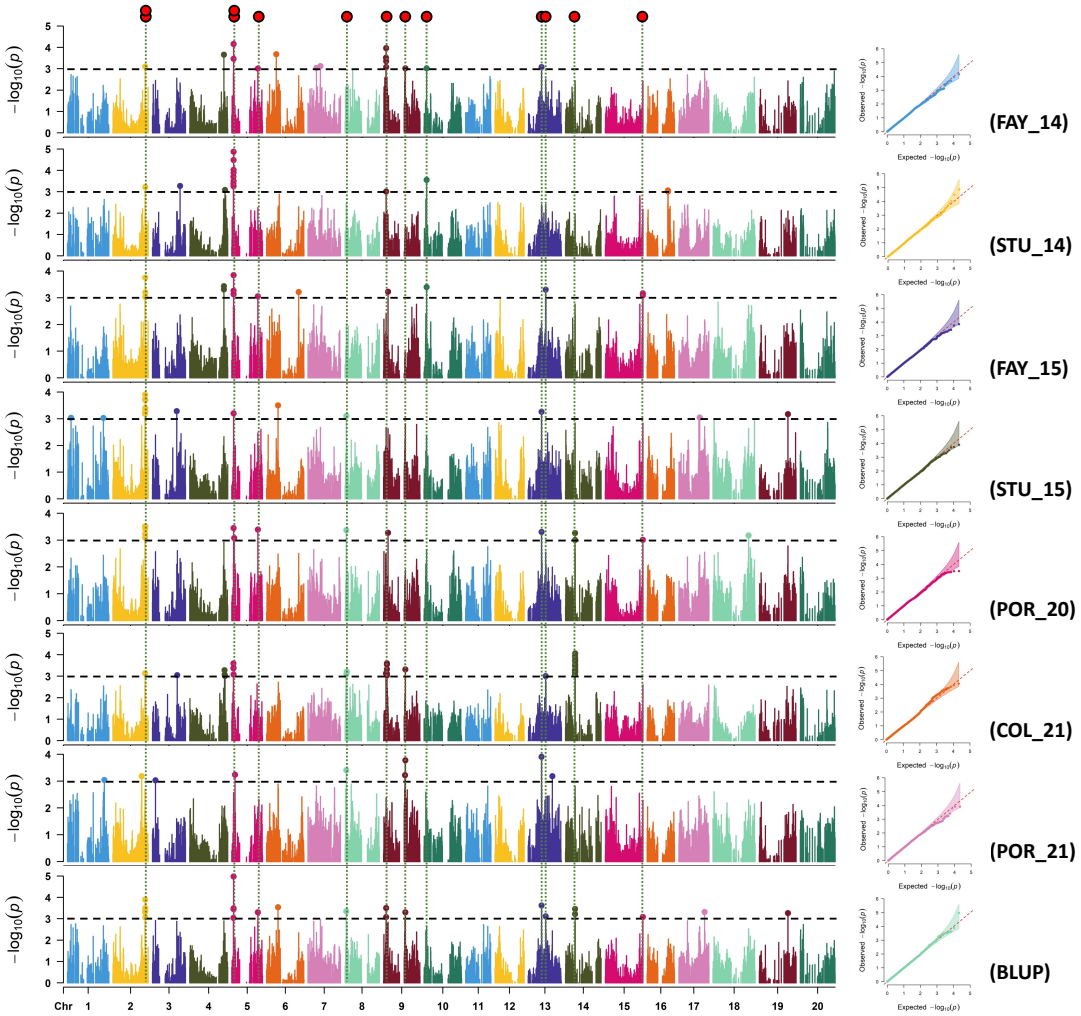


Figure 3. Manhattan and quantile-quantile plots of stachyose for FAY_14 (Fayetteville in 2014), STU_14 (Stuttgart in 2014), FAY_15 (Fayetteville in 2015), STU_15 (Stuttgart in 2015), POR_20 (Portageville in 2020), COL_21 (Columbia in 2021), POR_21 (Portageville in 2021), and BLUP (across seven environments, an additional environment). SNP markers detected in three or more environments were considered significant SNPs using GWAS threshold $-\log_{10}(P) > 3$ (red dots on the top).

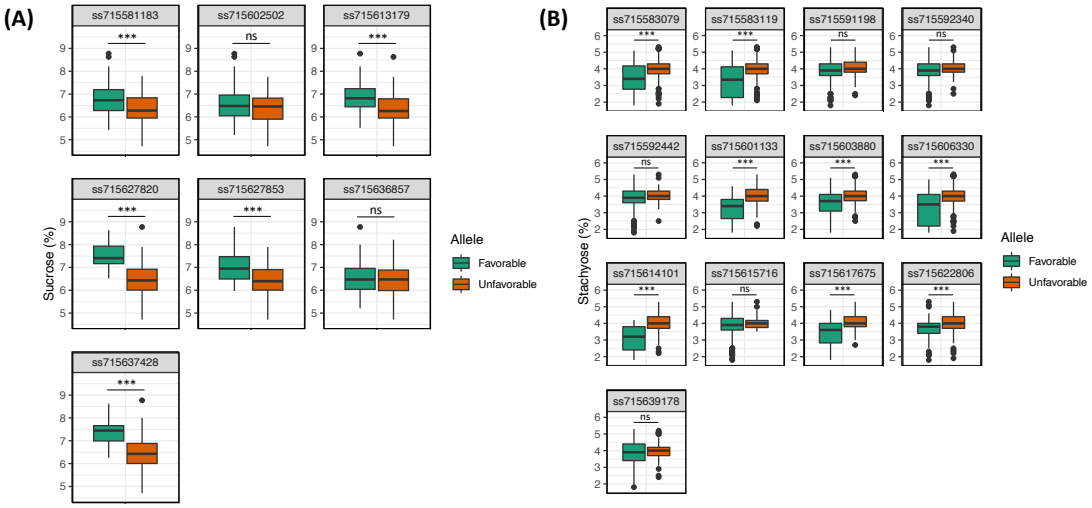


Figure 4. Allelic effect of significant SNPs for sucrose (A) and stachyose (B) content using mean values of favorable and unfavorable alleles across seven environments. The allele conferring higher sucrose content was designated a favorable allele, while the allele conferring lower stachyose content was designated a favorable allele. ns, no significant; *, significant level $p < 0.05$; **, significant level $p < 0.01$; and ***, significant level $p < 0.001$.

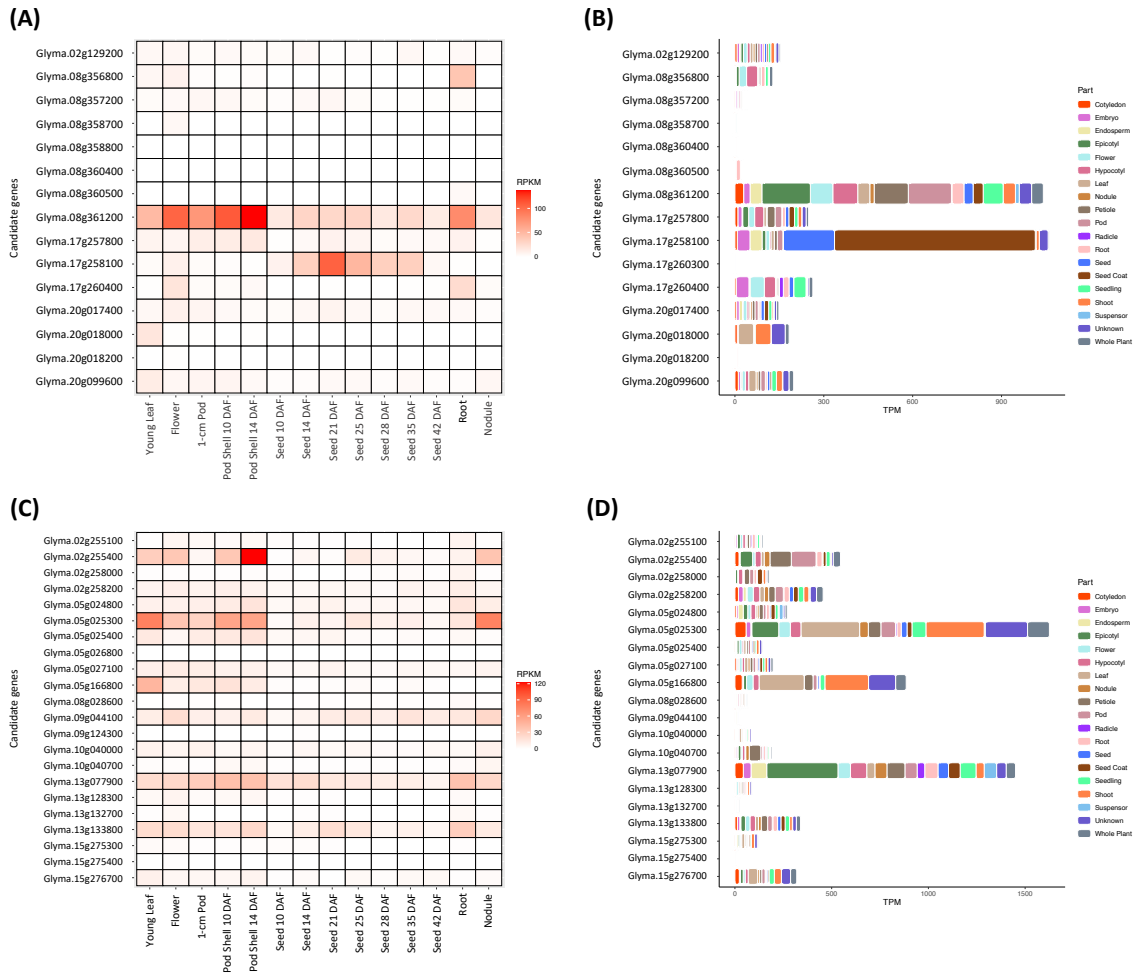


Figure 5. Gene expression levels of the potential candidate genes in 14 different tissues and 19 parts of soybean plant for sucrose content (A and B, respectively) and stachyose content (C and D, respectively). The raw gene expression counts were normalized using a Reads Per Kilobase of transcript per Million mapped reads (RPKM) and a Transcripts Per Million (TPM) method.

CHAPTER 3

Discovery of novel variants for sucrose content via QTL mapping and whole-genome sequencing analysis in soybean [*Glycine max* (L.) Merr.]

ABSTRACT

Sucrose in soybean [*Glycine max* (L.) Merr.] is nutritionally valuable for animal feed efficiency and plays a vital role in the sweetness of soy products. In this study, two recombinant inbred line (RIL) populations derived from the same donor parent, PI 506593, were used for quantitative trait loci (QTL) analysis employing BARCSoySNP6K Illumina Infinium BeadChips for seed sucrose content across three growing environments. Whole-genome sequencing (WGS) analysis based on two reference genome assemblies, Wm82.a2.v1 and Wm82.a5.v1, was implemented to discover novel variants in candidate genes. The analysis identified 11 sucrose-related QTL regions on chromosomes (Chrs.) 4, 5, 6, 8, 10, and 13. Among them, four QTL (*qSUC_08.1*, *qSUC_08.2*, *qSUC_08.3*, and *qSUC_08.4*) were clustered on Chr. 8 based on Wm82.a2.v1. We considered this cluster a major QTL region conferring high sucrose content in PI 506593. Within the major QTL region, 44 candidate genes with unique variants present only in PI 506593 were found based on WGS analysis using Wm82.a2.v1 genome assembly. A subsequent WGS analysis using Wm82.a5.v1 genome assembly confirmed that only 31 candidate genes overlapped between Wm82.a2.v1 and Wm82.a5.v1. GO enrichment analysis and functional annotations were further investigated for the 31 candidate genes. Three genes, *Glyma.08g294300*, *Glyma.08g297000*, and *Glyma.08g300400,* were identified to be responsible for sucrose metabolism-related enzymes, including galactosyltransferase, the regulation of the carbohydrate metabolic process, and glycosyltransferase, respectively. The novel variants of candidate genes identified in this study will provide new genetic resources to improve seed sucrose content for animal feed and human consumption.

INTRODUCTION

Soybean [*Glycine max* (L.) Merr.] is the leading protein meal source, accounting for 70.7% (244 million metric tons) of global protein meal consumption, followed by rapeseed (*Brassica napus* L.) (12.1%) in 2022 (http://soystats.com). Soybean meal is primarily used to feed non-ruminant animals, such as poultry and swine, as a main source of high-quality protein in feed formulations (Pope et al., 2023). Soybean meal meets three fundamental requirements in animal feeding programs, which are 1) having one or more essential nutrients; 2) being available to meet the demand of regular usage on a large scale; and 3) being cost-effective to use (Dozier and Hess, 2011). In addition to protein, soluble carbohydrates are another vital component in animal feed regarding metabolizable energy (ME) efficiency. Soybean seeds generally contain up to 15% soluble carbohydrates on a dry-weight basis (Feng et al., 2005). The three main components consisting of soluble carbohydrates are sucrose, raffinose, and stachyose, of which sucrose normally accounts for the largest portion of soluble carbohydrates in cultivated soybean seeds (Sui et al., 2020; Wang et al., 2023a).

Among soluble carbohydrates in soybean, sucrose is the only nutritionally beneficial component, which is readily digestible by non-ruminant animals and turns into the source of ME (Parsons et al., 2000; John, 2008; Jo et al., 2018; Wang et al., 2023a). Animal feeding programs recommend feed formulations with high sucrose content because sucrose has a significantly higher ME value (3,900 kcal/kg) than starch (2,918 - 3,396 kcal/kg) (John, 2008; Ostezan et al., 2023). Additionally, as the global vegetarian population has been increasing in recent decades, soy products have garnered popularity as a substitute for meat protein in vegetarian diets (Simeone et al., 2022; Wang et al., 2023b). Sucrose is a pivotal contributor to natural sweetness in soy food products, such as

edamame, miso, natto, tofu, and soymilk (Rosset et al., 2012; Sui et al., 2020; Wang et al., 2023a). In addition, the various health benefits of soy products have been well-studied in public health and functional food science, which include maintaining bone health and preventing heart disease, osteoporosis, and some cancers (Messina and Messina, 2010; Lee et al., 2019a; Chuang et al., 2021). Therefore, developing high sucrose soybean cultivars has been targeted in breeding programs for animal feed efficiency and natural sweetness to satisfy customers' preference for soy products.

Quantitative trait loci (QTL) mapping strategies have been widely used to identify genomic regions governing complex quantitative traits (Rani et al., 2023). Compared to protein and oil, sucrose content has garnered relatively less attention from soybean breeders. To date, 248 protein-related and 322 oil-related QTL have been reported from bi-parental QTL mapping studies. However, only 37 QTL for sucrose have been reported in Soybase (Soybase.org. Accessed on December 19, 2023). Maughan et al. (2000) first reported 17 sucrose-related QTL on chromosomes (Chrs.) 5, 7, 8, 13, 15, 19, and 20 using an interspecific cross between a large-seeded breeding line and a wild soybean (*G. soja*) line. Kim et al. (2005, 2006) identified a total of 17 sucrose-related QTL on Chrs. 2, 11, 12, 14, 15, 16, and 19 using two $F_{2:10}$ recombinant inbred line (RIL) populations sharing the same high sucrose donor line (Keunolkong). A major QTL for sucrose was identified on Chr. 11 in V99-5089, derived from a cross between V71-370 and PI 87013 (Saghai Maroof and Buss, 2008). Skoneczka et al. (2009) used two $F_{2:3}$ populations to identify another major QTL on Chr. 6 for high sucrose, which was derived from PI 200508. Later, three sucrose-related QTL on Chrs. 7, 11, and 20 were identified by Wang et al. (2014), of which QTL on Chr. 11 was the same QTL that Saghai Maroof and Buss (2008) previously identified.

Zeng et al. (2014) identified three sucrose-related QTL on Chrs. 5, 9, and 16. There have been more QTL mapping studies conducted for sucrose content in soybean seeds using different donor parents and larger marker datasets (Akond et al., 2015; Lee et al., 2016; Patil et al., 2018; Salari et al., 2021; Knizia et al., 2023; Liu et al., 2023).

However, only a few variants and genes for high sucrose content have been successfully employed in soybean breeding programs. A three-base pair deletion leading to an amino acid deletion (*rs2W331-*) and a missense mutation (*rs2T107I*) in raffinose synthase 2 (RS2, *Glyma.06g179200*, Wm82.a2.v1) coding regions were responsible for increased sucrose and reduced raffinose and stachyose content (Kerr and Sebastian, 2000; Dierking and Bilyeu, 2008, 2009; Skoneczka et al., 2009). Although mutations in RS3 (*Glyma.05g003900*, Wm82.a2.v1) (*rs3snp5*/*rs3snp6*) alone did not increase the sucrose content, the combination of the RS2 and RS3 mutations significantly increases sucrose content (Jo et al., 2018; Thapa et al., 2019; Hagely et al., 2020). A natural mutation in the D-myo-inositol 3-phosphate synthase 1 (MIPS1) gene (*Glyma.11g238800*, Wm82.a2.v1) derived from V99-5089 was associated with elevated sucrose content and significantly reduced raffinose and stachyose content in soybean (Saghai Maroof and Buss, 2008; Rosso et al., 2011).

Recently, a new high-quality reference genome sequence of Williams 82 (Wm82.a5.v1) was developed using the combination of PacBio HiFi sequencing and Bionano optical mapping (Garg et al., 2023). The Wm82.a5.v1 genome assembly conveys much smaller gaps and scaffold breaks compared to previous versions, such as Wm82.a1.v1.1 (Schmutz et al., 2010), Wm82.a2.v1 (Song et al., 2016), and Wm82.a4.v1 (Valliyodan et al., 2019). The previous reference genome assemblies contained 56,691 gaps

(42.29 Mbp), whereas Wm82.a5.v1 assembly is near gapless (only 14.25 Kbp gaps across the genome), of which ten chromosomes were completely reconstructed without any gaps (Garg et al., 2023). The most recent reference genome sequence data will provide a robust and accurate understanding of the genetic architecture underlying complex quantitative traits of interest.

In this study, a high sucrose germplasm, PI 506593, was crossed with two low sucrose breeding lines, S16-14161 and S16-11651, to develop two bi-parental RIL mapping populations. The populations were evaluated in three field environments to identify sucrose-related QTL using the BARCSoySNP6K Illumina Infinium BeadChips. Furthermore, we compared whole-genome sequence (WGS) data of parental lines aligned to Wm82.a2.v1 and Wm82.a5.v1 to identify candidate genes with unique SNP and Indel variants in PI 506593. This study provides valuable information on reliable QTL, candidate genes, and novel variants controlling sucrose content in soybean for improving seed compositional properties for animal feed and human consumption.

MATERIALS AND METHODS

Plant materials

PI 506593 was selected from the genome-wide association study panel used in Lee et al. (2023) based on the high and stable sucrose content across multiple environments. PI 506593 is a high-sucrose large-seeded (32g 100 seeds^{-1}) maturity group VI soybean line originating from Japan (https://www.ars-grin.gov/). Two high-yielding low-sucrose lines, S16-14161 and S16-11651 (Chen et al., 2022), developed by the University of Missouri – Fisher Delta Research, Extension, and Education Center (FDREEC), were used to cross with PI 506593.

The crosses of S16-14161 × PI 506593 and S16-11651 × PI 506593 were made in the summer of 2019 at the FDREEC in Portageville, MO, USA. The F_1 and F_2 generations were grown in the East Campus Plant Growth Facility, the University of Missouri, Columbia, MO, USA, and harvested in October 2020. The F_3 and F_4 generations were grown in the soybean winter nursery in Costa Rica until mid-June 2021. Seed generations were advanced using the single-seed descent method. A total of 140 and 207 $F_{4:5}$ RILs constituted two mapping populations, named P593_RIL1 (S16-14161 × PI 506593) and P593_RIL2 (S16-11651 × PI 506593), respectively, and were subsequently employed for QTL analysis.

Field experiments

The two RIL populations and the parental lines were evaluated for sucrose content in three field environments. It included the FDREEC, Portageville, MO, USA (36.42 °N 89.70 °W) in 2021 and 2022 (FDREEC_21 and FDREEC_22, respectively) and the Bradford Research and Extension Center (BREC), the University of Missouri, Columbia,

MO, USA (38.89 °N 92.19 °W) in 2022 (BREC_22). Ten seeds of each RIL and parental lines were planted in 75 cm wide rows in hill plots spaced 30 cm apart at the FDREEC and 60 cm apart at the BREC in a randomized complete block design (RCBD) with two replications. Seeds were harvested at maturity for sucrose quantification.

Phenotyping for sucrose content

The $F_{5:6}$ and $F_{6:7}$ RIL populations harvested in 2021 and 2022, respectively, were phenotyped for sucrose content. Ten intact soybean seeds (fully matured, spotless, and no damage) were sampled from each plot and used to quantify sucrose content using the established High-Performance Liquid Chromatography (HPLC) protocol described by Valliyodan et al. (2015) with minor modifications. Briefly, a powder sample was obtained by grinding seed samples with Thomas Wiley Mini-Mill (Arthur Thomas Co., Chadds Ford, PA, USA) fitted with a 20-mesh screen. The samples were lyophilized for two days using a Labconco freeze-dry system (Labconco, Kansas City, MO, USA). HPLC-grade water of 900 µL was added to 90.25 (± 0.15) mg of each lyophilized sample in a 2 mL centrifuge tube. After incubating tubes at 55°C with 200 rpm agitation for an hour, the tubes were vortexed for 30 seconds. After 20 minutes under room temperature, 900 µL HPLC-grade acetonitrile was added to each tube. Then, the suspension was centrifuged for 30 minutes at a 14.0 × 1000 min^{-1} × g speed. The supernatant was diluted five times using 65% HPLC-grade acetonitrile to prepare the final samples. The Agilent HPLC-ELSD (Evaporative Light Scattering Detection) 120 series (Agilent, Santa Clara, CA, USA), equipped with the Prevail Carbohydrate ES columns (5 µm 250 × 4.6 mm) and guard columns (7.5 × 4.6 mm) (Grace Davison Discovery Sciences, Deerfield, IL, USA) were used to quantify sucrose

content. The calibration curves were generated using the standard mixtures prepared in HPLC-grade water with 50, 100, 300, 500, and 1000 μg/mL concentrations.

Genotypic data

The genomic DNA of the $F_{5:6}$ RIL populations and parental lines were isolated from the youngest trifoliate leaves using an established cetyltrimethylammonium bromide (CTAB) protocol with minor modifications as previously described (Vuong et al., 2010). Marker genotyping was performed in the Soybean Genomics and Improvement Laboratory, USDA-ARS, Beltsville, MD, using the BARCSoySNP6K Illumina Infinium BeadChips (Illumina Inc., San Diego, CA, USA) (Song et al., 2020). The SNP alleles were called using the Genotyping Module 2.0 in GenomeStudio software (https://www.illumina.com/). The physical positions of SNP markers were aligned to Wm82.a2.v1. The WGS data of parental lines were generated using the Illumina NovaSeq PE150 platform at a depth of 30× by Novogene Corporation Inc. (Sacramento, CA, USA). Sequencing libraries with an insert size of ~350 bp were constructed incorporating the standard protocol.

Whole-genome sequence data processing

The raw Illumina reads were trimmed using Trimmomatic version 0.39 with the following parameters: ILLUMINACLIP:~/Trimmomatic-0.39/adapters/TruSeq_Adapters.fa:2:30:10 LEADING:10 TRAILING:10 SLIDINGWINDOW:5:20 MINLEN:50 (Bolger et al., 2014). The clean data were aligned to two reference genome assemblies, Wm82.a2.v1 and Wm82.a5.v1, using the BWA-MEM command (Li and Durbin, 2009). The genome assemblies and annotation files of

Wm82.a2.v1 and Wm82.a5.v1 were obtained from Soybase (https://www.soybase.org/) and Garg et al. (2023), respectively. The SAM files were converted to sorted BAM files using the SortSam (parameter: SORT_ORDER=coordinate) and MarkDuplicates (parameter: VALIDATION_STRINGENCY=SILENT SORTING_COLLECTION_SIZE_RATIO=0.15) modules of the Genome Analysis Toolkit (GATK) (McKenna et al., 2010). Variant calling was performed using the HaplotypeCaller module. SNP variants and Indels were called using the SelectVariants module in GATK. Variant filtering was performed using the VariantFiltration module to apply stringent filtering to sites with a quality depth (QD) < 2, fisher strand (FS) > 60, mapping quality (MQ) < 40, strand odds ratio (SOR) > 4, mapping quality rank sum (MQRankSum) < -12.5, and read position rank sum (ReadPosRankSum) < -8 for SNPs, and QD < 2, FS > 200, and SOR > 10 for Indels. Variants were annotated using the ANNOVAR tool (Wang et al., 2010). The liftover file containing corresponding gene IDs and their physical positions between Wm82.a2.v1 and Wm82.a5.v1 was graciously provided by Dr. Aamir Khan, the co-inventor of Wm82.a5.v1 (Garg et al., 2023).

Genetic map construction and QTL analysis

A genetic linkage map was constructed for each population using JoinMap version 5.0 software (Van Ooijen, 2018). A logarithm of the odds (LOD) score from 2 to 15 was used to cluster linkage groups. Regression mapping was set for the mapping algorithm, and map distances were calculated using Kosambi's mapping function. QTL analysis was carried out using MapQTL version 6.0 software based on a genetic linkage map, SNP alleles, and phenotypic data (Van Ooijen, 2009). Successive procedures, including interval

mapping (IM), automatic cofactor selection (ACS), and multi-QTL method (MQM), were performed for QTL analysis. The genome-wide LOD threshold was determined using a 1,000 permutation test at the 0.05 probability level of significance, of which the LOD threshold was 3.2 for both populations in this study. The QTL identified in this study were named according to the following rule: *qSUC*_chromosome.QTL number.

Candidate gene identification

Among all genes harbored in QTL identified, candidate genes with nonsynonymous SNP and Indel mutations present only in PI 506593 were initially screened using WGS data aligned to Wm82.a2.v1. The WGS data aligned to Wm82.a5.v1 were further used to confirm the presence of mutations in candidate genes. Subsequently, the candidate gene list was narrowed down based on the functional annotations related to sucrose metabolism using Phytozome and Gene Ontology (GO) enrichment analysis. Nucleotide and amino sequences of the final candidate genes of parental lines were compared based on Wm82.a5.v1 using the Jalview version 2.11.3.2 (Waterhouse et al., 2009).

Statistical analysis

The broad-sense heritability (H^2) was calculated using the *H2cal* function in the *inti* R package (https://rdrr.io/cran/inti/src/R/H2cal.R). Analysis of variance (ANOVA) was conducted using the PROC MIXED procedure of SAS software version 9.4 (SAS Institute, Cary, NC, USA). Genotype was used as a fixed effect, and Environment, Genotype × Environment interaction, and replication were used as random effects.

RESULTS

Phenotypic variations of sucrose content in two RIL populations

The sucrose contents of two RIL populations (P593_RIL1 and P593_RIL2) across three environments are shown in Table 1. In the P593_RIL1, sucrose content showed transgressive segregations with a normal distribution (Shapiro-Wilk value from 0.988 to 0.992) in all environments (Table 1). PI 506593 consistently showed higher sucrose content (8.2%) than S16-14161 (5.5%) across environments. P593_RIL1 had the highest sucrose content in BREC_22 (8.3%) and the lowest in FDREEC_21 (6.9%). The coefficient of variation (CV) ranged from 10.4% to 12.9%, and the broad-sense heritability was 0.83. Similarly, the sucrose content of P593_RIL2 showed normal distributions (Shapiro-Wilk value from 0.975 to 0.996) and transgressive segregations in all environments (Table 1). Also, PI 506593 consistently showed higher sucrose content (8.2%) than S16-11651 (6.5%) in P593_RIL2 across environments. The sucrose content of P593_RIL2 also varied across environments, following a similar trend as P593_RIL1, with the highest mean sucrose content observed in BREC_22 (7.8%) and the lowest in FDREEC_21 (6.6%). The CV ranged from 10.0% to 12.6%, and the broad-sense heritability was 0.84. The ANOVA showed that the effects of all factors were significant except the genotype × environment interaction effect in P593_RIL1 (Supplementary Table 1).

Genetic linkage map for P593_RIL1 and P593_RIL2

Among 6,000 markers from the BARCSoySNP6K BeadChips, 1,440 (24%) and 600 (10%) markers were polymorphic between the two parents of the P593_RIL1 (S16-14161 × PI 506593) and P593_RIL2 (S16-11651 × PI 506593), respectively (Table 2). The

polymorphic markers were anchored into 20 chromosomes of the integrated genetic map with a total genetic distance of 2,422.4 cM in the P593_RIL1 and 2,695.2 cM in the P593_RIL2. In the P593_RIL1, the total distance of each chromosome ranged from 74.9 cM (Chr. 7) to 187.5 cM (Chr. 13), and the genetic interval between two neighboring markers averaged from 1.3 cM to 2.6 cM, with an average of 1.7 cM. In the P593_RIL2, the total distance of each chromosome ranged from 54.2 cM (Chr. 7) to 217.1 cM (Chr. 2), and the average genetic interval between two neighboring markers varied from 2.0 cM to 15.5 cM, with an average of 5.2 cM.

Identification of QTL associated with sucrose content

Analyses of the two mapping populations identified 11 sucrose-related QTL on Chrs. 4, 5, 6, 8, 10, and 13 (Table 3). In the P593_RIL1, six QTL were identified on Chrs. 6 (*qSUC_06.1* and *qSUC_06.2*), 8 (*qSUC_08.1* and *qSUC_08.2*), 10 (*qSUC_10.1*), and 13 (*qSUC_13.1*) and explained 9.1 – 14.9% of the phenotypic variation (Table 3). Among them, *qSUC_08.1* and *qSUC_06.1* had the highest LOD values of 5.9 and 5.3 in the FDREEC_21 and FDREEC_22, respectively. Only one QTL (*qSUC_06.2*) was detected with a LOD value of 4.1 in BREC_22. The additive effects of each QTL ranged from -0.4 to -0.3, indicating that high sucrose-related alleles were inherited from the donor parent, PI 506593.

In the P593_RIL2, five QTL were mapped on Chrs. 4 (*qSUC_04.1*), 5 (*qSUC_05.1*), 8 (*qSUC_08.3* and *qSUC_08.4*), and 13 (*qSUC_13.2*) and explained 5.5 – 19.7% of the phenotypic variation (Table 3). Among them, *qSUC_08.3* and *qSUC_08.4* showed the highest LOD values of 12.1 and 3.8 in FDREEC_21 and BREC_22, respectively, while

qSUC_08.4 was the only QTL detected in FDREEC_22 with a LOD value of 5.3. Similar to the P593_RIL1, the additive effects of each QTL varied from -0.4 to -0.2, indicating that PI 506593 positively contributed to an increase in sucrose content in the P593_RIL2. QTL analysis of two populations consistently pinpointed a QTL region on Chr. 8 harboring four QTL (*qSUC_08.1*, *qSUC_08.2*, *qSUC_08.3*, and *qSUC_08.4*) at the physical interval of 40,597,410 – 42,861,364 bp across all environments (except BREC_22 in the P593_RIL1) (Table 3). We considered this region a major QTL conferring high sucrose content in PI 506593 for candidate gene identification.

Candidate gene identification

Based on Wm82.a2.v1, 175 genes were identified in the major QTL region on Chr. 8. Among these, 44 genes carried 85 nonsynonymous SNPs and 11 Indels that were only present in PI 506593 (Supplementary Tables 2 and 3). The subsequent WGS analysis based on Wm82.a5.v1 revealed that 31 candidate genes overlapped between two genome assemblies, while 13 genes on Wm82.a2.v1 were not found on Wm82.a5.v1. (Figure 1; Supplementary Tables 4 and 5). Based on functional annotations obtained from GO enrichment analysis and Phytozome, three candidate genes, *Glyma.08g294300*, *Glyma.08g297000*, and *Glyma.08g300400*, were responsible for galactosyltransferase, the regulation of carbohydrate metabolic process, and glycosyltransferase activity, respectively, which are closely related to sucrose metabolism (Figure 1; Table 4).

Novel variants of candidate genes

Based on Wm82.a2.v1, the physical position of *Glyma.08g294300* was from 40,953,540 to 40,957,782 bp (4,243 bp length) on Chr. 8 with seven exons (Figure 2A; Table 4). The sequence alignment between parental lines showed a nonsynonymous SNP substitution from cytosine to adenine (C to A) located at 40,956,342 bp on exon 5 in PI 506593 on Wm82.a2.v1 (Figure 2A). After liftover, the corresponding gene, *Gm_Wm82_23224*, was at the physical position of 43,407,555 – 43,415,567 bp (8,013 bp length) with ten exons on Wm82.a5.v1 (Figure 2B; Table 4). Although there was a difference in the number of exons, the identical nonsynonymous SNP substitution (C to A) was still present at 43,410,052 bp on exon 5 of PI 506593 on Wm82.a5.v1 (Figure 2B). This SNP mutation (C1618A) led to one amino acid change at the position of 540 from leucine to isoleucine (L540I) (Figure 2C).

G*lyma.08g297000* was located at the physical position of 41,326,371 – 41,329,183 bp (2,813 bp length) on Chr. 8 with eight exons on Wm82.a2.v1 (Figure 3A; Table 4). A nonsynonymous SNP substitution from adenine to cytosine (A to C) was identified at 41,328,845 bp on exon 8 in PI 506593 on Wm82.a2.v1 (Figure 3A). The corresponding gene on Wm82.a5.v1 was *Gm_Wm82_23251*, located at the physical position of 43,779,198 – 43,781,761 bp (2,564 bp length) on Chr. 8 (Table 4). The number of exons were same between *Glyma.08g297000* and *Gm_Wm82_23251* (Figure 3B). Furthermore, the identical nonsynonymous SNP mutation (A to C) was present at 43,781,623 bp on exon 8 in PI 506593 on Wm82.a5.v1. This SNP mutation (A1066C) led to one amino acid change at the position of 362 from glutamic acid to aspartic acid (E362D) (Figure 3C).

Based on Wm82.a2.v1, *Glyma.08g300400* was located at the physical position of 41,830,891 – 41,843,406 bp (12,516 bp length) on Chr. 8 with seven exons included

(Figure 4A; Table 4). A nonsynonymous SNP substitution from guanine to adenine (G to A) was identified at 41,838,405 bp on exon 2 in PI 506593 on Wm82.a2.v1 (Figure 4A). The Wm82.a5.v1 corresponding gene was *Gm_Wm82_23285*, located at 44,278,415 – 44,283,458 bp (5,044 bp length) with nine exons included (Figure 4B; Table 4). Although the number of exons differed, the identical nonsynonymous SNP substitution (G to A) was found at 44,292,981 bp in exon 2 in Wm82.a5.v1 (Figure 4B). This SNP mutation (G334A) led to one amino acid change at the position of 112 from valine to methionine (V112M) (Figure 4C).

DISCUSSION

Similar to other seed composition traits in soybean, sucrose content fluctuates due to various environmental factors, such as temperature, precipitation, and air and soil humidity, that affect plant growth, flowering period, and seed development (Tarumingkeng and Coto, 2003; Jo et al., 2018; Lee et al., 2023; Wang et al., 2023a). In this study, the average sucrose content was the highest in BREC_22 and the lowest in FDREEC_21 (Table 1). Although many studies elucidated a reversed correlation between temperature and sucrose content (the lower temperature induces the higher sucrose) during pod-filling stages, our study did not thoroughly follow this trend. From late September to October, however, the average temperature of BREC_22 was the lowest at 15.2°C, while FDREEC_21 had the highest temperature of 17.8°C based on the weather data from nearby weather stations (https://www.wunderground.com/). This observation suggests that the sucrose content of soybean may be more sensitive to lower temperatures during the late reproductive stages (R6 – R7) than during the early pod-filling stages (R3 – R5). Soil moisture is another important environmental factor affecting the sucrose content, although the correlation remains to be clarified (Wijewardana et al., 2019; Du et al., 2020; Wang et al., 2021). We utilized the dew point temperature to determine a correlation between soil moisture and sucrose content in this study. The dew point temperature provides an actual amount of water vapor in the air and positively correlates with soil moisture content (Leelamanie, 2010). In this study, the sucrose content was negatively correlated with the dew point temperature during the pod-filling stages, indicating that lower soil moisture levels induced more sucrose content. This result was consistent with previous studies

showing the negative correlation between soil moisture and sucrose content in soybean (Du et al., 2020; Wang et al., 2023a).

A total of 11 QTL on Chrs. 4, 5, 6, 8, 10, and 13 were identified in two populations across three environments (Table 3). Four QTL (*qSUC_08.1*, *qSUC_08.2*, *qSUC_08.3*, and *qSUC_08.4*) on Chr. 8 were detected across environments except BREC_22 in P593_RIL1, of which *qSUC_08.4* was detected in both FDREEC_22 and BREC_22 (Table 3). Since two populations shared the same donor parent, a genomic region harboring these four QTL on Chr. 8 was considered a major QTL region (the physical interval of 40,597,410 – 42,861,364 bp, Wm82.a2.v1) that underlies the high sucrose trait in PI 506593. This major QTL region was located near the QTL (43,602,528 - 43,670,249 bp) previously identified by Lee et al. (2016) using a population derived from a cross of Keunolkong × Iksan 10 (Supplementary Table 6). However, they did not overlap. Other minor QTL identified on Chrs. 4, 5, 6, 10, and 13 seemed unstable and inconsistent across environments (Table 3). Two QTL, *qSUC_06.1* and *qSUC_06.2*, were only identified on Chr. 6 in the P593_RIL1 at the physical intervals of 16,133,204 – 21,786,787 bp and 21,786,787 – 46,805,251 bp, respectively (Table 3). However, due to the lack of polymorphic markers on Chr. 6, the two QTL constituted more than 50% of the entire chromosome, which spans approximately 30 Mbp. Therefore, most of the QTL previously identified on Chr. 6 were located within this region (Skoneczka et al., 2009; Patil et al., 2018; Liu et al., 2023) (Supplementary Table 6). A minor QTL on Chr. 13 (*qSUC_13.1*) identified in P593_RIL1 was located at the physical interval of 25,268,844 – 25,927,261 bp, which was close to the QTL that Maughan et al. (2000) previously identified at the physical interval of 26,196,486 – 28,912,864 bp

('Sucrose 1-5' in Soybase) (Supplementary Table 6). The other minor QTL on Chrs. 4, 5, 10, and 13 were novel (Supplementary Table 6).

In this study, the whole-genome sequencing data of the three parental lines (PI 506593, S16-14161, and S16-11651) constituting the two mapping populations were analyzed for SNP and Indel variants discovery. This genomic information allowed us to narrow down the candidate genes in sucrose-related QTL identified by a conventional bi-parental mapping analysis. Furthermore, this study was the first to employ the latest reference genome, Wm82.a5.v1 (Garg et al., 2023). To date, most of novel genomic findings in soybean research were identified and reported based on the Wm82.a2.v1 assembly version. However, in the WGS analysis, some variants identified in Wm82.a2.v1 were not found on Wm82.a5.v1 (Supplementary Tables 4 and 5). This is because most of the gaps and scaffold breaks in Wm82.a2.v1 were completed in Wm82.a5.v1. Thus, although it was an early stage of the application of Wm82.a5.v1, the comparison in the sequence of parental lines between Wm82.a2.v1 and Wm82.a5.v1 provided robust results in identifying the novel variants of candidate genes for sucrose content.

Although only variants and candidate genes in the major QTL on Chr. 8 were fully disclosed in this study, all novel variants and candidate genes in other QTL on Chrs. 4, 5, 6, 10, and 13 were simultaneously investigated and described in Supplementary Tables 3, 4, 5, and 6. Within the major QTL on Chr. 8 of PI 506593, three nonsynonymous SNP substitutions, C1618A, A1066C, and G334A, were identified on *Glyma.08g294300*, *Glyma.08g297000*, and *Glyma.08g300400*, respectively, which are closely related to sucrose metabolism (Table 4). *Glyma.08g294300* encodes galactosyltransferase and is homologous to the *GATL4 – 6* genes, sharing 58 – 60% amino acid similarities in

Arabidopsis thaliana. *GALT* genes also encode the galactosyltransferase family protein in *Arabidopsis thaliana*. Galactosyltransferase plays a key role in transferring galactose and regulating carbon partitioning between sucrose and raffinose in soybean seeds (Saravitz et al., 1987; Singer et al., 2023). *Glyma.08g297000* is responsible for the regulation of the carbohydrate metabolic process (GO:0006109) and response to sucrose (GO:0009744) (Table 4). Its homologous gene in *Arabidopsis thaliana* is *WRI1* (*WRINKLED 1*), which encodes the transcription factor vital for the control of storage compound biosynthesis in *Arabidopsis thaliana*. Also, *WRI1* is known to trigger the expression of some sugar-responsive genes (Kong et al., 2019). *Glyma.08g300400* is responsible for transferase activity, transferring glycosyl groups (GO:0016757) and establishing α-1,6-glycosidic bonds. Its homologous genes in *Arabidopsis thaliana* were *AT2G02910* (73% amino acid similarity) and *AT1G34550* (59% amino acid similarity), of which both genes are known for glycosyltransferase activity based on TAIR (https://www.arabidopsis.org/). Given roles in carbohydrate metabolism, the novel variants in three candidate genes may block the conversion from sucrose to raffinose by disturbing the establishment of an α-1,6-glycosidic bond between sucrose and galactose molecules.

Lastly, we also investigated the occurrence of these novel variants in other soybean germplasms with high sucrose content, namely PI 506937 and PI 561292B, as well as in the high-sucrose experimental lines, KB10-23 1681 0, S17-16989, and R07-2000, which carry known mutations of *rs2W331-/rs3snp6*, *rs2W331-/rs3snp6*, and *mips1*, respectively. Interestingly, the novel variants identified in this study were also found in PI 506937 and PI 561292B, while these variants were not present in KB10-23 1681 0, S17-16989, and R07-2000 (Supplementary Figure 1). It elucidates that the novel variants may be

widespread natural mutations conferring high sucrose content in soybean seeds. Future investigations, including gene transformation (overexpression or knockout) and molecular marker development, will be required to characterize the candidate genes and novel variants associated with sucrose content in soybean seeds.

CONCLUSION

QTL analysis using two mapping populations derived from the same donor parent, PI 506593, identified a major QTL on Chr. 8 controlling sucrose content in soybean seeds. The candidate gene list was narrowed down by identifying SNP and Indel variants of candidate genes using WGS data aligned to Wm82.a2.v1 and Wm82.a5.v1. Three novel SNP variants were found in three candidate genes, *Glyma.08g294300*, *Glyma.08g297000*, and *Glyma.08g300400* in PI 506593. The discovery of novel SNP variants will enhance breeding efficiency in high-sucrose soybean cultivars for animal feed and human consumption.

REFERENCES

Akond, M., Liu, S., Kantartzi, S.K., Meksem, K., Bellaloui, N., Lightfoot, D.A., et al. (2015). Quantitative Trait Loci Underlying Seed Sugars Content in "MD96-5722" by "Spencer" Recombinant Inbred Line Population of Soybean. *Food and Nutrition Sciences*, 6, 964-973. https://doi.org/10.4236/fns.2015.611100

Bolger, A.M., Lohse, M., and Usadel, B. (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics*, 30(15), 2114-2120. https://doi.org/10.1093/bioinformatics/btu170

Chen, P., Ali, M.L., Shannon, G., Vieira, C.C., Lee, D., Crisel, M., et al. (2022). Registration of 'S16-11651C', a conventional soybean cultivar with high yield, resistance to multiple diseases, and broad adaptation. *Journal of Plant Registration*, 16(2), 329-340. https://doi.org/10.1002/plr2.20224

Chuang, T.L., Lin, C.H., and Wang, Y.F. (2021). Effects of vegetarian diet on bone mineral density. *Tzu Chi Medical Journal*, 33(2), 128-134. https://doi.org/10.4103/tcmj.tcmj_84_20

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

Dierking, E.C., and 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

Dozier, W.A., and Hess, J.B. (2011). Soybean meal quality and analytical techniques. In H. El-Shemy (Ed.), *Soybean and Nutrition* (pp. 111-124). IntechOpen. https://doi.org/10.5772/1008

Du, Y., Zhao, Q., Chen, L., Yao, X., Zhang, H., Wu, J., et al. (2020). Effect of Drought Stress during Soybean R2–R6 Growth Stages on Sucrose Metabolism in Leaf and Seed. *Molecular Sciences*, 21(2), 618. https://doi.org/10.3390/ijms21020618

Fehr, W.R. (1987). Breeding methods for cultivar development. In J.R. Wilcox (Ed.), *Soybeans: Improvement, production, and uses* (2nd ed., pp. 249–293). ASA, CSSA, and SSSA.

Feng, C.B., Morsy, M., Giannoccaro, E., Zhang, B., and Chen, P. (2005). Soybean seed sugar content and quantitative trait loci mapping. In C.J. Li, et al. (Eds.), *Plant nutrition for food security, human health, and environmental protection* (pp. 438-439). Tsinghua University Press.

Garg, V., Khan, A.W., Fengler, K., Llaca, V., Yuan, Y., Vuong, T.D., et al. (2023). Near-gapless genome assemblies of Williams 82 and Lee cultivars for accelerating global soybean research. *The Plant Genome*, 16(4), e20382. https://doi.org/10.1002/tpg2.20382

Hagely, K.B., Jo, H., Kim, J.H., Hudson, K.A., and 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

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

John, P.G.K. (2008). Sugar syrup: the new energy feed for poultry. *World Poultry Science*, 24, 12-13. Available at: https://edepot.wur.nl/9105. Accessed 14 December 2023.

Kerr, P.S., and Sebastian, S.A. (2000). Soybean products with improved carbohydrate composition and soybean plants. United States Patent 6,147,193.

Kim, H.K., Kang, S.T., and 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, H.K., Kang, S.T., Cho, J.H., Choung, M.G., and 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

Knizia, D., Bellaloui, N., Yuan, J., Lakhssassi, N., Anil, E., Vuong, T., et al. (2023). Quantitative Trait Loci and Candidate Genes That Control Seed Sugars Contents in the Soybean 'Forrest' by 'Williams 82' Recombinant Inbred Line Population. *Plants*, 12(19), 3498. https://doi.org/10.3390/plants12193498

Kong, Q., Yuan, L., and Ma, W. (2019). WRINKLED1, a "Master Regulator" in Transcriptional Control of Plant Oil Biosynthesis. *Plants*, 8(7), 238. https://doi.org/10.3390/plants8070238

Lee, D., Kulkarni, K.P., Kim, B., Seok, Y.M., Song, J.T., and Lee, J.D. (2019). Comparative assessment of quality characteristics of Chungkookjang made from soybean seeds differing in oleic acid concentration. *Journal of Functional Foods*, 52, 529-536. https://doi.org/10.1016/j.jff.2018.10.016

Lee, D., Lara, L., Moseley, D., Vuong, T.D., Shannon, G., Xu, D., and Nguyen, H.T. (2023). Novel genetic resources associated with sucrose and stachyose content through genome-wide association study in soybean [*Glycine max* (L.) Merr.]. *Frontiers in Plant Science*, 14, 1294659. https://doi.org/10.3389/fpls.2023.1294659

Lee, D., Vuong T.D., Shannon, G., Shi, H., Rainey, K.M., Nguyen, H.T. (2023). Environmental stability and genetic effect of soybeans differing in mutant allele combinations between rs and mips1 genes for soluble carbohydrate profiles. *Crop Science*, 63(6), 3326-3337. https://doi.org/10.1002/csc2.21094

Lee, J.S., Kim, S.M., and Kang, S. (2016). Fine mapping of quantitative trait loci for sucrose and oligosaccharide contents in soybean [*Glycine max* (L.) Merr.] using 180 K Axiom SoyaSNP genotyping platform. *Euphytica*, 208, 195-203. https://doi.org/10.1007/s10681-015-1622-x

Leelamanie, D.A.L. (2010). Changes in Soil Water Content with Ambient Relative Humidity in Relation to the Organic Matter and Clay. *Tropical Agricultural Research and Extension*, 13(1), 6-10. http://dx.doi.org/10.4038/tare.v13i1.3130

Li, H., and Durbin, R. (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics*, 25(14), 1754-1760. https://doi.org/10.1093/bioinformatics/btp324

Liu, C., Chen, H., Yu, Q., Gu, H., Li, Y., Tu, B., et al. (2023). Identification of quantitative trait loci and candidate genes for seed sucrose and soluble sugar concentrations in soybean. *Crop Science*, 63, 2976-2992. https://doi.org/10.1002/csc2.21080

Maughan, P.J., Saghai Maroof, M.A., and 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

McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernytsky, A., et al. (2010). The genome analysis toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Research*, 20(9), 1297-1303. https://doi.org/10.1101/gr.107524.110

Messina, M., and Messina, V. (2010). The role of soy in vegetarian diets. *Nutrients*, 2(8), 855-888. https://doi.org/10.3390/nu2080855

Ostezan, A., Prenger, E.M., Rosso, L., Zhang, B., Stupar, R.M., Glenn, T., et al. (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

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

Patil, G., Vuong, T.D., Kale, S., Valliyodan, B., Deshmukh, R., Zhu, C., et al. (2018). Dissecting genomic hotspots underlying seed protein, oil, and sucrose content in an interspecific mapping population of soybean using high density linkage mapping. *Plant Biotechnology Journal*, 16, 1939-1953. https://doi.org/10.1111/pbi.12929

Pope, M., Borg, B., Boyd, R.D., Holzgraefe, D., Rush, C., and Sifri, M. (2023). Qualifying the value of soybean meal in poultry and swine diets. *Journal of Applied Poultry Research*, 32(2), 100337. https://doi.org/10.1016/j.japr.2023.100337

Rani, K., Kumar, M., Razzaq, A., Ajay, B.C., Kona, P., Bera, S.K., et al. (2023). Recent advances in molecular marker technology for QTL mapping in plants. In S.H. Wani, D. Wang, and G.P. Singh (Eds.), *QTL Mapping in Crop Improvement: Present Progress and Future Perspectives* (pp. 1-15). Academic Press.

Rosset, M., Prudencio, S.H., and 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(6), 531-538. https://doi.org/10.1177/1082013211433076

Rosso, M.L., Burleson, S.A., Maupin, L.M., and 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

Saghai Maroof, M.A., and Buss, G.R. (2008). Low phytic acid, low stachyose, high sucrose soybean lines. U.S. Patent 12/033,830, field Feb.19, 2008.

Salari, M.W., Ongom, P.O., Thapa, R., Nguyen, H.T., Vuong, T.D., Rainey, K.M. (2021). Mapping QTL controlling soybean seed sucrose and oligosaccharides in a single family of soybean nested association mapping (SoyNAM) population. *Plant Breeding*, 140(1), 110-122. https://doi.org/10.1111/pbr.12883

Saravitz, D.M., Pharr, D.M., and Carter, T.E. (1987). Galactinol Synthase Activity and Soluble Sugars in Developing Seeds of Four Soybean Genotypes. *Plant Physiology*, 83(1), 185-189. https://doi.org/10.1104/pp.83.1.185

Schmutz, J., Cannon, S., Schlueter, J., Ma, J., Mitros, T., Nelson, W., et al. (2010). Genome sequence of the palaeopolyploid soybean. *Nature*, 463, 178-183. https://doi.org/10.1038/nature08670

Simeone, G., Bergamini, M., Verga, M.C., Cuomo, B., D'Antonio, G., Iacono, I.D., et al. (2022). Do Vegetarian Diets Provide Adequate Nutrient Intake during Complementary Feeding? A Systematic Review. *Nutrients*, 14(17), 3591. https://doi.org/10.3390/nu14173591

Singer, W.M., Lee, Y., Shea, Z., Vieira, C.C., Lee, D., Li, X., et al. (2023). Soybean genetics, genomics, and breeding for improving nutritional value and reducing antinutritional traits in food and feed. *The Plant Genome*, 16(4), e20415. https://doi.org/10.1002/tpg2.20415

Skoneczka, J.A., Saghai Maroof, M.A., Shang, C., and Buss, G.R. (2009). Identification of Candidate Gene Mutation Associated With Low Stachyose Phenotype in Soybean Line PI200508. *Crop Science*, 49, 247-255. https://doi.org/10.2135/cropsci2008.07.0403

Song, Q., Jenkins, J., Jia, G., Hyten, D.L., Pantalone, V., Jackson, S.A., et al. (2016). Construction of high resolution genetic linkage maps to improve the soybean genome sequence assembly Glyma1.01. *BMC Genomics*, 17, 33. https://doi.org/10.1186/s12864-015-2344-0

Song, Q., Yan, L., Quigley, C., Fickus, E., Wei, H., Chen, L., et al. (2020). Soybean BARCSoySNP6K: An assay for soybean genetics and breeding research. *The Plant Journal*, 104(3), 800–811. https://doi.org/10.1111/tpj.14960

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

Tarumingkeng, R.C., and Coto, Z. (2003). Effects of drought stress on growth and yield of soybean. *Kisman Science Philosophy*, 702, 798-807.

Valliyodan, B., Cannon, S.B., Bayer, P.E., Shu, S., Brown, A.V., Ren, L., et al. (2019). Construction and comparison of three reference-quality genome assemblies for soybean. *The Plant Journal*, 100(5), 1066-1082. https://doi.org/10.1111/tpj.14500

Valliyodan, B., Shi, H., and Nguyen, H.T. (2015). A simple analytical method for high-throughput screening of major sugars from soybean by normal-phase HPLC with evaporative light scattering detection. *Chromatography Research International*, 2015, 8. https://doi.org/10.1155/2015/757649

Van Ooijen, J.W. (2009). MapQTL® 6, Software for the mapping of quantitative trait loci in experimental populations of diploid species. *Kyazma BV, Wageningen, Netherlands*, 64.

Van Ooijen, J.W. (2018). JoinMap® 5, Software for the calculation of genetic linkage maps in experimental populations of diploid species. *Kyazma BV, Wageningen, Netherlands.*

Vuong, T.D., Sleper, D.A., Shannon, J.G., and Nguyen, H.T. (2010). Novel quantitative trait loci for broad-based resistance to soybean cyst nematode (*Heterodera glycines Ichinohe*) in soybean PI 567516C. *Theoretical and Applied Genetics*, 121, 1253-1266. https://doi.org/10.1007/s00122-010-1385-7

Wang, K., Li, M., and Hakonarson, H. (2010). ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. *Nucleic Acids Research*, 38(16), e164. https://doi.org/10.1093/nar/gkq603

Wang, T., Masedunskas, A., Willett, W.C., and Fontana, L. (2023b). Vegetarian and vegan diets: benefits and drawbacks. *European Heart Journal*, 44(26), 3423-3439. https://doi.org/10.1093/eurheartj/ehad436

Wang, X., Wu, Z., Zhou, Q., Wang, X., Song, S., and Dong, S. (2021). Physiological Response of Soybean Plants to Water Deficit. *Frontiers in Plant Science*, 12, 809692. https://doi.org/10.3389/fpls.2021.809692

Wang, Y., Chen, P., and 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., Yu, D., Morota, G., Dhakal, K., Singer, W., Lord, N., et al. (2023a). 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

Waterhouse, A.M., Procter, J.B., Martin, D.M.A., Clamp, M., and Barton, G.J. (2009). Jalview Version 2 – a multiple sequence alignment editor and analysis workbench. *Bioinformatics*, 25(9), 1189-1191. https://doi.org/10.1093/bioinformatics/btp033

Wijewardana, C., Reddy, K.R., and Bellaloui, N. (2019). Soybean seed physiology, quality, and chemical composition under soil moisture stress. *Food Chemistry*, 278, 92-100. https://doi.org/10.1016/j.foodchem.2018.11.035

Zeng, A., Chen, P., Shi, A., Wang, D., Zhang, B., Orazaly, M., et al. (2014). Identification of quantitative trait loci for sucrose content in soybean seed. *Crop Science*, 54(2), 554-564. https://doi.org/10.2135/crops ci2013.01.0036

TABLES AND FIGURES

Table 1. Phenotypic variations of sucrose content in parents (P1 and P2) and two mapping populations across three environments.

Population[1]	Env[2]	P1	P2	Mean ± SD[3]	Range	Shapiro wilk (w)	Skewness	Kurtosis	Variance	CV (%)[4]	Heritability (H^2)[5]
P593_RIL1	FDREEC_21	4.9	7.8	6.9 ± 0.9	4.5 - 9.4	0.991	-0.01	3.02	0.89	12.9	0.83
	FDREEC_22	5.4	8.0	7.2 ± 0.9	5.3 - 10.1	0.988	0.30	3.34	0.75	10.4	
	BREC_22	6.3	8.8	8.3 ± 1.0	5.8 - 10.8	0.992	0.15	3.00	0.93	11.2	
	Mean	5.5	8.2	7.4 ± 0.8	5.1 - 9.4			-			
P593_RIL2	FDREEC_21	5.9	7.8	6.6 ± 0.9	4.4 - 9.1	0.99	0.07	2.61	0.83	12.6	0.84
	FDREEC_22	6.5	8.0	6.9 ± 0.9	3.8 - 9.5	0.996	-0.05	3.24	0.79	11.4	
	BREC_22	7.0	8.8	7.8 ± 0.9	6.1 - 10.8	0.975	0.54	3.1	0.78	10.0	
	Mean	6.5	8.2	7.1 ± 0.8	5.2 - 9.0			-			

[1]P593_RIL1, a bi-parental mapping population from a cross between S16-14161 (P1) and PI 506593 (P2); P593_RIL2, a bi-parental mapping population from a cross between S16-11651 (P1) and PI 506593 (P2). [2]FDREEC_21, Fisher Delta Research Education and Extension Center in 2021; FDREEC_22, Fisher Delta Research Education and Extension Center in 2022; BREC_22, Bradford Research Education Center in 2022. [3]SD, standard deviation. [4]CV, the coefficient of variation. [5]H^2, the broad sense of heritability.

Table 2. Distribution of SNP markers across 20 chromosomes.

Chr.	P593_RIL1			P593_RIL2		
	# Markers	Total distance (cM)	Average interval (cM)	# Markers	Total distance (cM)	Average interval (cM)
1	48	114.8	2.4	21	176.1	8.4
2	99	163.2	1.6	14	217.1	15.5
3	74	113.1	1.5	35	81.7	2.3
4	84	131.3	1.6	22	159.4	7.2
5	56	123.3	2.2	18	120.2	6.7
6	75	159.7	2.1	23	115.4	5
7	57	74.9	1.3	27	54.2	2
8	104	141.4	1.4	24	118.5	4.9
9	79	117.6	1.5	31	187.2	6
10	80	143.2	1.8	28	127.4	4.6
11	57	146.2	2.6	23	93	4
12	67	97.8	1.5	18	124.1	6.9
13	82	187.5	2.3	61	176.4	2.9
14	62	82	1.3	51	133.1	2.6
15	62	107.3	1.7	16	81	5.1
16	69	101.2	1.5	39	189.2	4.9
17	83	121.3	1.5	44	102.6	2.3
18	59	75.8	1.3	37	130	3.5
19	74	115.6	1.6	29	125.5	4.3
20	69	105.2	1.5	39	183.1	4.7
Total	**1440**	**2,422.4**	**1.7**	**600**	**2,695.2**	**5.2**

Table 3. Summary of QTL associated with sucrose content in two mapping populations across three environments.

Population	Environment	QTL[1]	Chr[2]	Marker interval (cM)	Physical interval (bp)	LOD[3]	PVE (%)[4]	Add[5]
P593_RIL1	FDREEC_21	*qSUC_08.1*	8	112.9 - 114.8	42,146,252 - 42,532,915	5.9	14	-0.4
	FDREEC_21	*qSUC_10.1*	10	49.9 - 51.5	37,502,435 - 37,995,354	4.5	10.4	-0.3
	FDREEC_21	*qSUC_13.1*	13	71.5 - 76.7	25,268,844 - 25,927,261	4	9.3	-0.3
	FDREEC_22	*qSUC_06.1*	6	115.2 - 124.0	16,133,204 - 21,786,787	5.3	14.9	-0.3
	FDREEC_22	*qSUC_08.2*	8	114.8 - 115.9	42,532,915 - 42,861,364	3.3	9.1	-0.3
	BREC_22	*qSUC_06.2*	6	124.0 - 131.4	21,786,787 - 46,805,251	4.1	13.9	-0.4
P593_RIL2	FDREEC_21	*qSUC_05.1*	5	77.5 - 86.1	34,272,637 - 35,101,291	5.2	7.8	-0.3
	FDREEC_21	*qSUC_08.3*	8	100.8 - 103.2	40,940,261 - 41,834,095	12.1	19.7	-0.4
	FDREEC_21	*qSUC_13.2*	13	116.9 - 117.1	32,525,645 - 32,570,888	3.7	5.5	-0.2
	FDREEC_22	*qSUC_08.4*	8	99.7 - 100.8	40,597,410 - 40,940,261	5.3	11.3	-0.3
	BREC_22	*qSUC_04.1*	4	106.0 - 125.4	46,793,849 - 48,002,103	3.4	6.3	-0.2
	BREC_22	*qSUC_08.4*	8	99.7 - 100.8	40,597,410 - 40,940,261	3.8	7.1	-0.3

[1]Quantitative Trait Loci. [2]Chromosome. [3]Logarithm of the odds. [4]Percentage variance explained. [5]Additive effect.

Table 4. Candidate genes in the major QTL region on Chr. 8 identified on Wm82.a2.v1 and Wm82.a5.v1 both.

Wm82.a2.v1			Wm82.a5.v1			Annotation
Gene ID	Position	Variant type[1]	Gene ID	Position	Variant type	
Glyma.08G293000	40,766,950 - 40,775,274	SNP	*Gm_Wm82_23206*	43,220,985 - 43,228,708	SNP	Sin3 family co-repressor
Glyma.08G293100	40,795,741 - 40,800,913	SNP	*Gm_Wm82_23209*	43,249,676 - 43,259,978	SNP	Sin3 family co-repressor
Glyma.08G293800	40,883,195 - 40,885,068	Indel	*Gm_Wm82_23219*	43,333,545 - 43,338,512	Indel	-[2]
Glyma.08G293900	40,887,462 - 40,910,754	SNP	*Gm_Wm82_23220*	43,341,530 - 43,364,303	SNP	Haloacid dehalogenase-like hydrolase
Glyma.08G294100	40,942,975 - 40,947,622	SNP	*Gm_Wm82_23222*	43,396,901 - 43,401,240	SNP	N-terminal protein myristoylation
Glyma.08G294300	40,953,540 - 40,957,782	SNP	*Gm_Wm82_23224*	43,407,555 - 43,415,567	SNP	Galactosyltransferase
Glyma.08G294800	40,993,962 - 41,001,629	SNP	*Gm_Wm82_23227*	43,447,708 - 43,455,285	SNP	Actin cross-linking protein
Glyma.08G295400	41,026,575 - 41,033,381	SNP/Indel	*Gm_Wm82_23230*	43,479,691 - 43,485,976	SNP/Indel	TIR domain
Glyma.08G295700	41,062,905 - 41,071,504	SNP	*Gm_Wm82_23234*	43,516,488 - 43,524,479	SNP	AMP deaminase
Glyma.08G295800	41,100,889 - 41,103,542	SNP	*Gm_Wm82_23237*	43,554,115 - 43,556,053	SNP	Homeobox associated leucine zipper
Glyma.08G296300	41,215,611 - 41,223,730	SNP	*Gm_Wm82_23243*	43,668,712 - 43,675,770	SNP	-
Glyma.08G296500	41,241,857 - 41,246,757	SNP	*Gm_Wm82_23245*	43,694,878 - 43,699,270	SNP	Transcription factor subunit Med10 of Mediator complex
Glyma.08G297000	41,326,371 - 41,329,183	SNP	*Gm_Wm82_23251*	43,779,198 - 43,781,761	SNP	Regulation of carbohydrate metabolic process
Glyma.08G297400	41,358,747 - 41,363,041	SNP	*Gm_Wm82_23255*	43,811,714 - 43,815,086	SNP	Protein kinase domain
Glyma.08G297500	41,378,414 - 41,383,832	SNP	*Gm_Wm82_23256*	43,831,982 - 43,836,271	SNP	IQ calmodulin-binding motif
Glyma.08G297700	41,412,445 - 41,415,775	SNP	*Gm_Wm82_23258*	43,864,913 - 43,868,558	SNP	-
Glyma.08G299500	41,767,501 - 41,769,958	SNP	*Gm_Wm82_23276*	44,211,989 - 44,214,205	SNP	Mitochondrial carrier protein
Glyma.08G299800	41,795,912 - 41,796,546	SNP/Indel	*Gm_Wm82_23279*	44,240,488 - 44,241,228	SNP	Chitinase
Glyma.08G300400	41,830,891 - 41,843,406	SNP	*Gm_Wm82_23285*	44,278,415 - 44,283,458	SNP	Glycosyltransferase activity

Glyma.08G300500	41,865,484 - 41,867,085	SNP	*Gm_Wm82_23287*	44,310,063 - 44,311,664	SNP	red, far-red light phototransduction
Glyma.08G301300	41,931,888 - 41,933,603	SNP	*Gm_Wm82_23293*	44,376,527 - 44,377,790	SNP	Ribosomal protein L34e
Glyma.08G301700	41,962,541 - 41,968,109	SNP/Indel	*Gm_Wm82_23298*	44,407,335 - 44,411,704	Indel	Homeobox domain
Glyma.08G302400	42,046,386 - 42,051,525	SNP	*Gm_Wm82_23302*	44,491,765 - 44,496,137	SNP	MIF4G domain
Glyma.08G303000	42,114,246 - 42,121,684	SNP/Indel	*Gm_Wm82_23307*	44,559,451 - 44,566,133	SNP/Indel	Helicase conserved C-terminal domain
Glyma.08G303500	42,150,074 - 42,151,723	SNP/Indel	*Gm_Wm82_23312*	44,595,014 - 44,596,723	SNP/Indel	Leucine-rich repeat (LRR) family protein
Glyma.08G304300	42,265,594 - 42,273,748	Indel	*Gm_Wm82_23320*	44,711,920 - 44,718,967	Indel	WD domain, G-beta repeat
Glyma.08G305600	42,374,266 - 42,376,352	SNP	*Gm_Wm82_23334*	44,821,303 - 44,823,057	SNP	Leucine-rich repeat (LRR) family protein
Glyma.08G307300	42,551,411 - 42,553,771	SNP	*Gm_Wm82_23351*	44,989,964 - 44,992,324	SNP	Protein kinase domain
Glyma.08G307400	42,556,212 - 42,559,048	SNP	*Gm_Wm82_23352*	44,994,964 - 44,997,445	SNP/Indel	Protein kinase domain
Glyma.08G307700	42,603,547 - 42,605,822	SNP	*Gm_Wm82_23357*	45,040,796 - 45,043,149	SNP	Ribosomal L18ae/LX protein domain
Glyma.08G309200	42,818,069 - 42,821,065	SNP	*Gm_Wm82_23372*	45,255,743 - 45,258,461	SNP	CDP-diacylglycerol biosynthesis I

[1]SNP, Single nucleotide polymorphism; Indel, Insertion and deletion. [2]-, Data not available

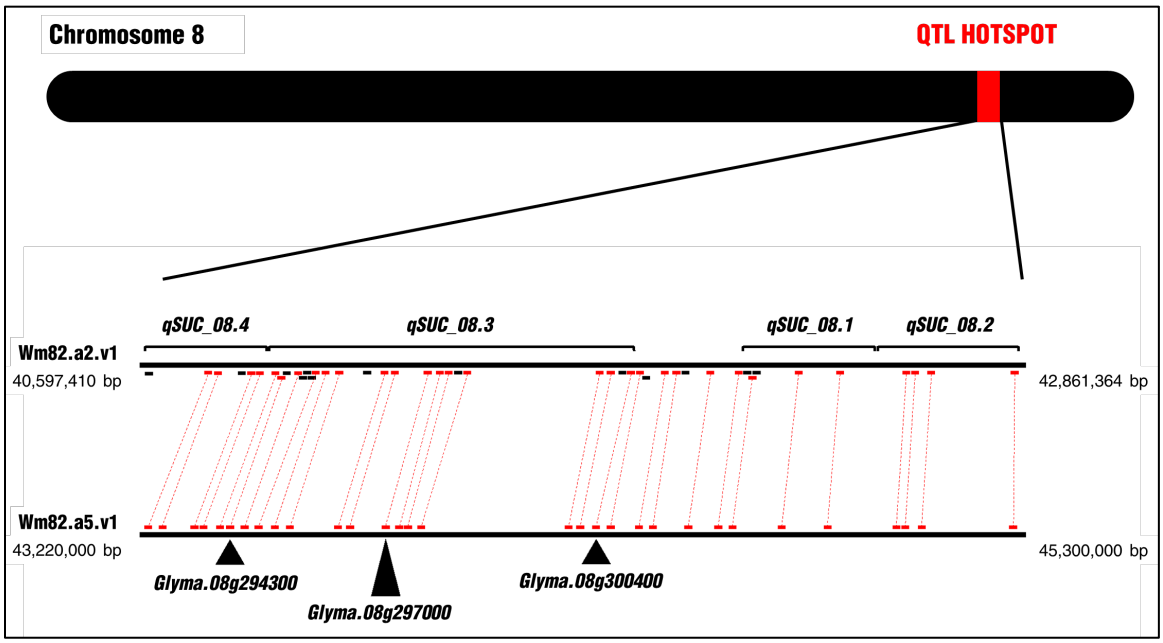


Figure 1. A comparison of the physical location of candidate genes between Wm82.a2.v1 and Wm82.a5.v1 genome assemblies and three candidate genes closely associated with sucrose metabolism. Candidate genes present in both Wm82.a2.v1 and Wm82.a5.v1 are marked in red, while those only present in Wm82.a2.v1 are marked in black.

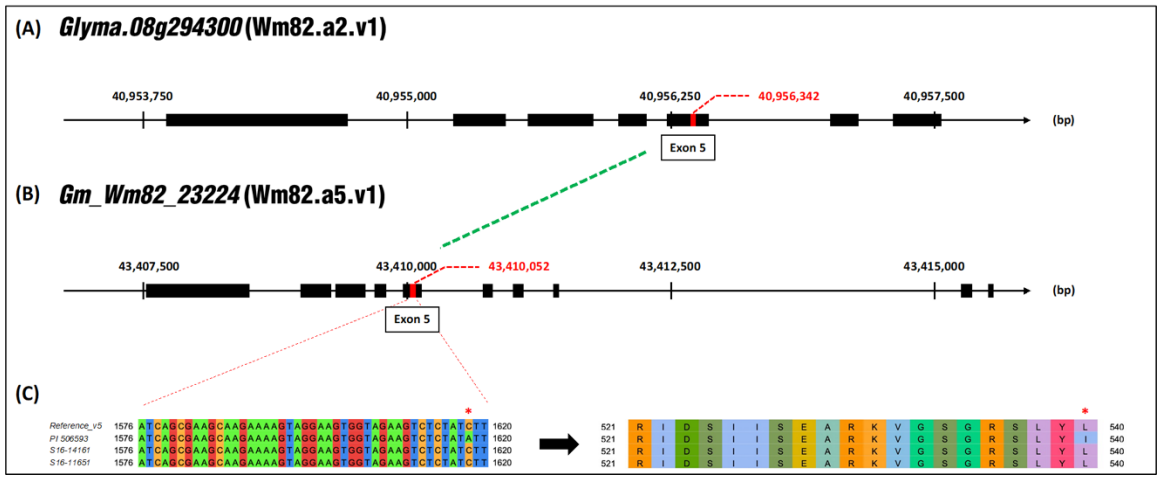


Figure 2. Location of the variant identified in *Glyma.08G294300* (Wm82.a2.v1) (A) and *Gm_Wm82_23224* (Wm82.a5.v1) (B) in PI 506593. Nucleotide and amino acid sequence alignment between parental lines based on Wm82.a5.v1 (C). The red asterisk (*) indicates the mutation in sequence.

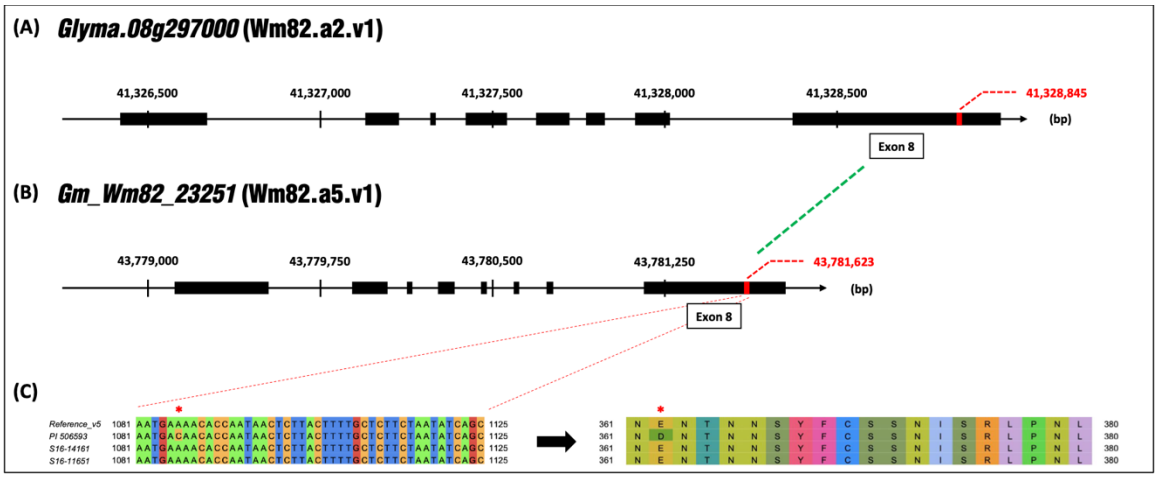


Figure 3. Location of the variant identified in *Glyma.08g297000* (Wm82.a2.v1) (A) and *Gm_Wm82_23251* (Wm82.a5.v1) (B) in PI 506593. Nucleotide and amino acid sequence alignment between parental lines based on Wm82.a5.v1 (C). The red asterisk (*) indicates the mutation in sequence.

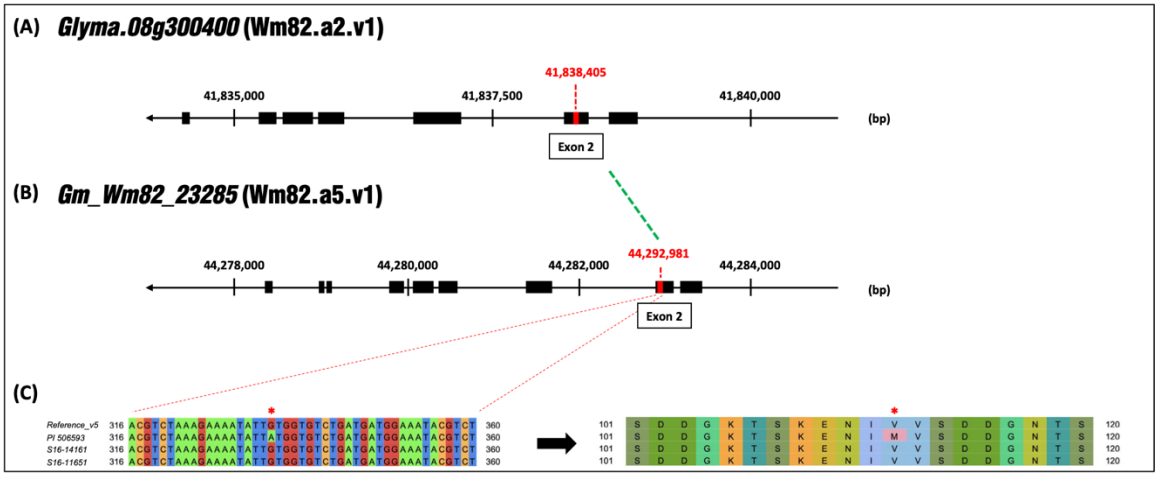


Figure 4. Location of the variant identified in *Glyma.08g300400* (Wm82.a2.v1) (A) and *Gm_Wm82_23285* (Wm82.a5.v1) (B) in PI 506593. Nucleotide and amino acid sequence alignment between parental lines based on Wm82.a5.v1 (C). The red asterisk (*) indicates the mutation in sequence.

CHAPTER 4

Environmental stability and genetic effect of soybeans differing in mutant allele combinations between *rs* and *mips1* genes for soluble carbohydrate profiles

Lee, D., Vuong, T.D., Shannon, G., Shi, H., Rainey, K.M., and Nguyen, H.T. (2023). Environmental stability and genetic effect of soybeans differing in mutant allele combinations between *rs* and *mips1* genes for soluble carbohydrate profiles. *Crop Science*, 63(6), 3326-3337. https://doi.org/10.1002/csc2.21094

ABSTRACT

Raffinose and stachyose (raffinose family oligosaccharides, RFOs) in soybean [*Glycine max* (L.) Merr.] are anti-nutritional carbohydrate components. Previous studies identified mutant alleles of raffinose synthase (*rs*) 2 and 3 and D-myo-inositol- 3-phosphate synthase 1 (*mips1*) to increase sucrose and reduce RFOs content. Although earlier studies investigated the effects of *rs2* and *rs3* mutations, a *mips1* mutation has not been tested yet. The objectives of this study were (1) to investigate the genetic effects and environmental stabilities of *rs2*, *rs3*, and *mips1* mutations across 18 environments and (2) to evaluate germination rates of genotypes carrying these mutations. Sucrose was significantly affected by all environmental factors (year, location, and planting dates). However, raffinose was not affected by planting dates, and stachyose was only affected by location. The *rs2* mutant allele (*rs2W331-*) and *mips1* showed the highest environmental stability for sucrose and raffinose, respectively. For stachyose, the wild-type check showed the highest environmental stability. The lines with *mips1* mutation showed significantly higher sucrose (11.1%) and lower raffinose (0.5%) and stachyose (0.6%) content than other genotypes across environments. The lines tested in field emergence tests showed no significant difference from the check. Lines with *mips1* and *rs2W331-*, respectively, showed significantly higher germination rates than the check in the accelerated aging test. In contrast, a line with *mips1* showed a significantly higher germination rate than the check in the standard germination test. Desirable carbohydrate allele combinations will facilitate the development of new cultivars with modified carbohydrate profiles that are beneficial to human food and animal feed.

INTRODUCTION

Due to its nutritional value and versatility, soybean [*Glycine max* (L.) Merr.] has become one of the economically important legume crops globally, producing 20% oil, 40% protein, and 15% soluble carbohydrates on a dry weight basis (Hsu et al., 1973; Liu, 1997). It is primarily used as a source of vegetable protein in foods and animal feed (Jo, 2016; Cunicelli et al., 2019). In 2020, soybean meal consumed in the animal feed industry was approximately 33 million metric tons, accounting for almost 74% of the total soybean meal production (45 million metric tons) in the United States (American Soybean Association, 2021). Poultry and swine accounted for approximately 60% (20 million metric tons) and 20% (6 million metric tons), respectively, of soybean meal consumed in the United States (American Soybean Association, 2021). Poultry and swine, however, are non-ruminant animals and have trouble digesting specific soluble carbohydrate components in soybean meal (Cunicelli et al., 2019; Jo et al., 2018, 2019).

Soluble carbohydrates in soybean seeds mainly comprise sucrose, raffinose, and stachyose; among these, raffinose and stachyose make up the raffinose family of oligosaccharides (RFOs) (Liu, 1997; Guillon and Champ, 2002; Karr-Lilienthal et al., 2005). Sucrose is a nutritionally beneficial component that provides metabolizable energy. In contrast, RFOs are indigestible in non-ruminant animals due to the lack of α-galactosidase in their monogastric systems to break down the α-1,6-glycosidic bonds of RFOs. Indigested RFOs pass through the upper gut and are fermented by anaerobic microbes in the lower gut, producing unwanted gastric discomforts, such as diarrhea and flatulence, resulting in a significant loss of energy efficiency in soybean meal (Coon et al., 1990; Le et al., 2020). The incorporation of genes for high sucrose and reduced RFOs is

considered an effective strategy to improve soybean meal quality in animal feed applications.

Significantly reduced RFOs and increased sucrose content were first discovered in the plant introduction (PI) 200508 in the USDA Soybean Germplasm Collection (Kerr and Sebastian, 2000). The desirable carbohydrate profile in PI 200508 was derived from a three-base pair deletion within the raffinose synthase 2 coding region (*RS2*, *Glyma.06g179200*, Wm82.a2.v1), resulting in one amino acid deletion (tryptophan) at a highly conserved position 331 (*rs2W331-*; Kerr and Sebastian, 2000; Dierking and Bilyeu, 2008; Skoneczka et al., 2009). A missense mutation in the *RS2* coding sequence was found in an ethyl-methanesulfonate (EMS) mutant line using reverse genetics by Targeting Induced Local Lesions IN Genomes (TILLING) (Dierking and Bilyeu, 2009). This mutation causes one amino acid change from threonine (T) to isoleucine (I) at position 107 (*rs2T107I*) (Dierking and Bilyeu, 2009). The extremely low RFOs content (approximately 1%) was only achieved by the combination of *RS2* and raffinose synthase 3 (*RS3*, *Glyma.05g003900*, Wm82.a2.v1) variant alleles (*rs3snp5*/*rs3snp6*) (Jo et al., 2018; Hagely et al., 2020). However, only variant alleles of *RS3* showed no significant effects on the carbohydrate content in soybean (Jo et al., 2018). Sebastian et al. (2000) developed a mutant line, LR33, with a single base pair polymorphism in a conserved region of D-myo-inositol-3-phosphate synthase 1 (MIPS1) gene (*Glyma.11g238800*, Wm82.a2.v1) induced by chemical mutagenesis. Although *mips1* conferred the desirable carbohydrate profile, *mips1* mutant soybean lines had significantly lower field emergence than non-mips1 mutant lines (Meis et al., 2003; Chappell et al., 2006). Later, Saghai Maroof and Buss (2008) developed a soybean line, 'V99-5089' carrying a novel natural variant of *MIPS1* with a

desirable carbohydrate profile (>11% sucrose and <1% RFOs) derived from a cross between ‘V71-370’ and ‘PI 87013’. Also, the nucleic acid sequence analysis showed the mutation in LR33 was not present in V99-5089 (Saghai Maroof and Buss, 2008).

Although genetic background primarily contributes to the phenotypic variation of carbohydrate content in soybean, environmental factors can also affect soybean carbohydrate content (Bellaloui et al., 2010; Jo et al., 2018, 2019). Bilyeu and Wiebold (2016) and Wolf et al. (1982) showed a distinct tendency for increased sucrose and reduced stachyose content in soybean under cooler temperatures during pod-filling stages. Also, late planting increased sucrose content since soybean planted late were generally exposed to cooler temperatures than those planted early (Jaureguy et al., 2013; Bilyeu and Wiebold, 2016). The objectives of this study were (1) to investigate the genetic and environmental effects on the stability of carbohydrate content among ten soybean genotypes carrying six different allele combinations (*rs2W331-*/*RS3*/*MIPS1*, *rs2T107I*/*RS3*/*MIPS1*, *rs2T107I*/*rs3snp6*/*MIPS1*, *rs2W331-*/*rs3snp6*/*MIPS1*, *RS2*/*RS3*/*mips1*, and *RS2*/*RS3*/*MIPS1*) across three geographically dispersed locations with two different planting dates in 2015–2017 and (2) to evaluate the germination rate of four soybean genotypes with different mutant alleles of *RS2*, *RS3*, and *MIPS1* using different test methods, including standard germination, accelerated aging, and field emergence tests.

MATERIALS AND METHODS

Multi-environmental study

Plant materials

Ten soybean lines were selected based on the combination of mutant alleles, including *rs2W331-*, *rs2T107I*, *rs3snp6*, and *mips1*, that affect carbohydrate content (Table 1). These lines were classified into six genotypic groups representing different mutant allele combinations. Two soybean lines developed by Dr. Grover Shannon at the University of Missouri – Fisher Delta Research, Extension, and Education Center (FDREEC), CR13-395 and CR13-397, carry one mutant allele of *RS2* (*rs2W331-*/*RS3*/*MIPS1*, group 1). The soybean lines with 'KB' designation were developed by the University of Missouri (Dr. Kristin Bilyeu). KB10-22#1548 b and KB10-22#1568 b carry one mutant allele of *RS2* (*rs2T107I*/*RS3*/*MIPS1*, group 2). KB10-22#1600 0 and KB10-22#1608 0 carry one mutant allele of *RS2* and one mutant allele of *RS3* (*rs2T107I*/*rs3snp6*/*MIPS1*, group 3). KB10-23#1681 0 carries one mutant allele of *RS2* and one mutant allele of *RS3* (*rs2W331-*/*rs3snp6*/*MIPS1*, group 4). R07-2000 and R07-2001 were developed by the University of Arkansas (Dr. Pengyin Chen) and carry one mutant allele of *MIPS1* (*RS2*/*RS3*/*mips1*, group 5). A public cultivar, LD06-7620, carries none of the mutant alleles (*RS2*/*RS3*/*MIPS1*, group 6) and was used as a check.

Locations and field experiments

For three years, 2015–2017, replicated experiments were conducted at three locations: Arkansas Agricultural Experiment Station, Fayetteville, AR (36.06 °N 94.16 °W), Agronomy Center for Research and Education, Lafayette, IN (40.43 °N 86.92 °W), and

FDREEC, Portageville, MO (36.42 °N 89.70 °W). About ten seeds of each soybean line were planted in hill plots spaced 61 cm apart between hills. A randomized complete block design (RCBD) was implemented with two replications on two planting dates (early and late plantings at a 3-week interval) at each location. The monthly average temperatures in July, August, September, and October where soybean is under seed development stages (R3–R6; Fehr et al., 1971) were collected from the nearest weather station of each location at https://www.wunderground.com/.

Carbohydrate phenotyping

Sample and standard preparation

Soluble carbohydrate content was determined by Agilent HPLC-ELSD 1200 series, equipped with the Prevail Carbohydrate ES columns (5 µm 250 × 4.6 mm) and guard columns (7.5 × 4.6 mm) (Valliyodan et al., 2015). Around ten seeds harvested from the hill plot were ground using Thomas Wiley Mini-Mill (Arthur Thomas Co., Philadelphia, PA, USA) fitted with a 20-mesh screen. The ground powder was lyophilized for two days in a Labconco freeze-dry system (Labconco, USA). Dried powder of 90.25 (± 0.15) mg was weighed and mixed with 900 µL HPLC-grade water in 2-mL centrifuge vials. The vials were incubated at 55°C with 200 rpm agitation for 1 hour, followed by a high-speed vortex for 30 seconds. After cooling at room temperature for 20 minutes, 900 µL HPLC-grade acetonitrile was added to each vial. The suspension was centrifuged for 30 minutes at 14.0 × 1000 min^{-1} × g. The supernatant was diluted five times with 65% HPLC-grade acetonitrile for HPLC analysis. Carbohydrate standards were prepared in HPLC-grade water with the final concentrations of 50, 100, 300, 500, and 1000 µg/mL.

HPLC instrumental method

Two types of mobile phases were prepared: mobile phase A was pure water, and mobile phase B was acetonitrile: acetone mixture of 75:25 (v/v). The flow rate and gradient of phases A and B were optimized for the complete separation of the soluble carbohydrates within 20 minutes. Column temperature remained at 35°C throughout the gradient program. The detector temperature was isothermal at 55°C. The nebulizer pressure was 3.4 bar. Ultra-purity-grade nitrogen (grade 5.0) was used as the nebulizer gas. The sample injection volume was 5 μL.

Germination and field emergence tests

Plant materials

Standard germination (SG) test, accelerated aging (AA) test, and field emergence test were conducted to evaluate the germination rate of four soybean genotypes, including R07-2000 (*RS2*/*RS3*/*mips1*), KB07-15 (*rs2W331-*/*RS3*/*MIPS1*), 15045-13 (*RS2*/*rs3-Tilling*/*MIPS1*), and Williams 82 (*RS2*/*RS3*/*MIPS1*). The testing seeds were harvested at the Bradford Research and Extension Center (BREC) in 2021 and stored in the seed storage room (4°C temperature and 36% relative humidity).

Standard germination (SG) test

The SG test was conducted by Missouri Crop Improvement Association (MCIA) (Columbia, MO 65201, USA) (https://moseed.org/) using an established standard protocol. Briefly, 50 seeds of each genotype were planted on a paper substratum as a rolled towel test in eight replications in a germination chamber with a temperature of 25°C and a relative

humidity of 75%–80% under the continuous light duration (AOSA, 2002). After around seven days, the final germination rate was calculated by averaging the germination rates of all replications.

Accelerated aging (AA) test

The AA test was also conducted by MCIA. First, seeds were rapidly aged under a temperature of 41°C and a relative humidity of 90%–95% under continuous light in a water-jacket aging chamber for 72 hours (AOSA, 2002). After aging, seeds were planted on a paper substratum as a rolled towel test with four replications of 50 seeds of each genotype in a germination chamber with a temperature of 25°C and a relative humidity of 75%–80% under continuous light duration. After around seven days, a final germination rate was calculated by averaging the germination rates of all replications.

Field emergence test

Field emergence test was conducted at the BREC, Columbia, MO and the FDREEC, Portageville, MO. One hundred seeds of each genotype were planted using a RCBD in a 2 m single row with two replications. Seeds were spaced 2 cm long. The final emergence rate was determined by averaging replications after ten days.

Statistical analysis

Analysis of variance and carbohydrate comparisons

An analysis of variance (ANOVA) was conducted using '*Genotype*' (individual soybean lines) and '*gGroup*' (genotypic groups) as a fixed effect, respectively, and the

'*Year*', '*Location*', '*Planting date*', and the interactions as random effects. The least significant difference (LSD) at *p*-value of 0.05 was used to determine the significant difference among means. The ANOVA was conducted using PROC GLM function in the SAS software version 9.4 (SAS Institute, Cary, NC, 2012). The significant difference among means of genotypes and genotypic groups was determined using PROC ANOVA function in the SAS software version 9.4.

Environmental stability analysis

Eighteen environment codes (E1–E18) were defined based on the combination of years (3), locations (3), and planting dates (2) to compare the environmental stability of genotypes for sucrose, raffinose, and stachyose content (Table 2). The environmental stability of soluble carbohydrate content was analyzed by using a method of 'Additive Main effect and Multiplicative Interaction' (AMMI) that integrates ANOVA and principal component analysis (PCA) into a unified approach (Gauch, 1988). The '*Genotype*' and '*Environment*' were fitted as additive main effects using a two-way ANOVA, followed by PCA using '*GEI* (*Genotype and Environment Interaction*)' as a multiplicative effect (Ajay et al., 2019; Singamsetti et al., 2021). AMMI biplots were constructed based on the main effect of means and the first principal component axis (PCA1). The equation used for AMMI (Singamsetti et al., 2021) is as follows:

$$Y_{ij} = \mu + g_i + e_j + \sum_{k=1}^{n} \delta_k \alpha_{ik} \gamma_{jk} + \theta_{ij}$$

where, Y_{ij} is the mean of each carbohydrate component of the '*Genotype*' i (i = 1, 2,…, G) in the environment j (j = 1, 2,…, E); μ is the general mean, g_i is the i^{th} genotypic effect;

e_j is the j^{th} environmental effect; δ_k is the eigenvalue of the PCA axis k; α_{ik} and γ_{jk} are the i^{th} genotype j^{th} environment PCA scores for the PCA axis k; θ_{ij} is the residual. The AMMI analysis was conducted using the software R (R Core Team, 2021) and package '*metan'* (Olivoto and Lúcio, 2020).

RESULT

Multi-environmental study

ANOVA and carbohydrate variation across the environments

The effects of *Genotype*, *Year*, *Location*, and *Planting date*, and the interactions on each carbohydrate component were determined by ANOVA (Table 3). Among all sources of variation, the effects of *Genotype*, *Location*, and the interactions with *Genotype* were significant for all carbohydrate components. The *Year* effect was only significant for sucrose and raffinose, while the *Planting date* effect was only significant for sucrose ($p < 0.001$). The effect of *Year* × *Location* × *Planting date* was only significant for sucrose and stachyose.

Similarly, the effects of *gGroup*, Location, and the interactions with *gGroup* (except for Gg × Y × L) were significant for all carbohydrate components (Table 4). The *Year* effect was only significant for sucrose and raffinose, while the *Planting date* had only a significant effect on sucrose ($p < 0.01$). The interaction, *Year* × *Location* × *Planting date*, had a significant effect on sucrose and stachyose ($p < 0.05$).

The phenotypic variation of carbohydrate content across three years, three locations, and two planting dates along with monthly average temperatures during pod-filling stages (July – October) were analyzed (Table 5). The year 2016 had relatively warmer temperatures than the years 2015 and 2017 during the pod-filling stages. The sucrose content in 2015 and 2017 was significantly higher than in 2016. Although the stachyose content was not statistically different over the years, the stachyose content in 2015 and 2017 was slightly lower than in 2016. The northernmost location, Lafayette, IN, showed the lowest average temperatures during the pod-filling stages, while Fayetteville, AR, was

the warmest location in 2015 – 2017. The sucrose content in Lafayette, IN was significantly higher than in Fayetteville, AR, and Portageville, MO. The warmest location, Fayetteville, AR, had the lowest sucrose content. The lowest stachyose content was observed in Fayetteville, AR. *Planting date* had a significant effect on sucrose only. The late planting date significantly increased sucrose content, but raffinose and stachyose content were not significantly different between early and late planting dates. Furthermore, the overall phenotypic patterns of each carbohydrate component of genotypes in different years, locations, and planting dates are shown in Figure 1. The most distinct pattern was the higher sucrose contents in the lowest temperature location, Lafayette, IN. Interestingly, higher stachyose contents were mostly observed in Portageville, MO, across three years, although the average temperature was intermediate. The phenotypic variation of raffinose was smaller than that of sucrose and stachyose.

The average soluble carbohydrate contents of ten soybean genotypes across three locations and three years were compared (Figure 2). The sucrose contents of ten genotypes ranged from 5.2% to 11.1% on a dry weight basis (Figure 2A). The two soybean lines in Group 5 (*RS2*/*RS3*/*mips1*) had significantly higher sucrose contents than the other soybean lines. The sucrose content of line 7 in Group 4 (*rs2W331-*/*rs3snp6*/*MIPS1*) was significantly higher than line 1 in Group 1 (*rs2W331-*/*RS3*/*MIPS1*), line 3 in Group 2 (*rs2T107I*/*RS3*/*MIPS1*), and line 10 in Group 6 (*RS2*/*RS3*/*MIPS1*). Also, line 7 had a higher sucrose content than line 2 in Group 1 (*rs2W331-*/*RS3*/*MIPS1*), line 4 in Group 2 (*rs2T107I*/*RS3*/*MIPS1*), and two lines in Group 3 (*rs2T107I*/*rs3snp6*/*MIPS1*), although it was not significantly different. This indicates that *rs2* and *rs3* combination increased sucrose content more effectively than a single mutation. Furthermore, the allelic effect of

rs3snp6 for sucrose was significant, but it was also genotype-dependent with *rs2W331-* or *rs2T107I* allele. The wild-type reference, line 10, showed the lowest sucrose content.

The raffinose contents of 10 genotypes ranged from 0.53% to 1.09% on a dry weight basis (Figure 2B). Group 5 (*RS2*/*RS3*/*mips1*) had a significantly lower raffinose content than other soybean lines, followed by the wild-type reference. Among *rs2* and *rs3* mutant lines, line 1 (*rs2W331-*/*RS3*/*MIPS1*) had the highest raffinose content, but it was not significantly different than the raffinose content of other genotypes except for line 4 in Group 2 (*rs2T107I*/*RS3*/*MIPS1*).

The stachyose contents of ten genotypes ranged from 0.6% to 4.3% on a dry weight basis (Figure 2C). The stachyose contents of two lines in Group 5 (*RS2*/*RS3*/*mips1*) were significantly lower than the stachyose contents of other soybean lines, followed by line 2 in Group 1 (*rs2W331-*/*RS3*/*MIPS1*). The single allele effect of *rs2W331-* was significantly higher than *rs2T107I* to reduce stachyose content. Additional *rs3snp6* allele significantly reduced stachyose contents in soybean lines with *rs2T107I*. Also, the combinations of *RS2* and *RS3* mutant alleles, *rs2T107I*/*rs3snp6*/*MIPS1* and *rs2W331-*/*rs3snp6*/*MIPS1*, were not significantly different from each other in stachyose content, although *rs2W331-*/*rs3snp6*/*MIPS1* showed a slightly lower stachyose content.

Environmental stability and interaction analysis

The environmental stability and interaction of soluble carbohydrate content across the 18 testing environments (E1 – E18) were visualized by AMMI biplots (Figure 3). For the interpretation of AMMI biplot, the displacements along the abscissa and ordinate were used to indicate differences in the main effects and interactions, respectively. For example,

for sucrose, line 2 (*rs2W331-*/*RS3*/*MIPS1*) had a PCA1 score close to zero, indicating the highest environmental stability, while line 7 (*rs2W331-*/*rs3snp6*/*MIPS1*) and line 9 (*RS2*/*RS3*/*mips1*) were far from the PCA1 score of zero, indicating the least environmental stability (Figure 3A). Although line 9 had the least environmental stability, it showed the highest sucrose content among the genotypes across the environments. E4 (Lafayette, IN, late planting, 2015) had the longest vector, indicating the strongest environmental interaction with sucrose. On the other hand, E5 (Portageville, MO, early planting, 2015) had the shortest vector, indicating the least environmental interaction with sucrose. For raffinose, line 8 (*RS2*/*RS3*/*mips1*) showed the highest environmental stability, while line 7 (*rs2W331-*/*rs3snp6*/*MIPS1*) was the least environmental stability across the environments (Figure 3B). Lines 8 and 9 (*RS2*/*RS3*/*mips1*) had the lowest raffinose contents among the genotypes across the environments. E4 (Lafayette, IN, late planting, 2015) had the strongest environmental interaction with raffinose, while E8 (Fayetteville, AR, late planting, 2016) had the least environmental interaction. The most environmentally stable genotype for stachyose was line 10 (*RS2*/*RS3*/*MIPS1*), while the most environmentally sensitive genotype was line 7 (*rs2W331-*/*rs3snp6*/*MIPS1*) (Figure 3C). Lines 8 and 9 (*RS2*/*RS3*/*mips1*) had significantly lower stachyose contents among genotypes across the environments. E4 (Lafayette, IN, late planting, 2015) had the strongest environmental effect on stachyose. In contrast, E7 (Fayetteville, AR, early planting, 2016) had the least environmental effect on stachyose.

The mean soluble carbohydrate contents of ten genotypes and six genotypic groups across 18 environments were visualized in heatmaps (Figure 4). Lines 8 and 9 in Group 5 (*RS2*/*RS3*/*mips1*) consistently showed higher sucrose content than other genotypes across

18 environments (Figures 4A and 4D). The environments of Lafayette, IN, including E3, E4, E10, E15, and E16, were favorable for higher sucrose content among the 18 environments. These results confirmed that lower temperatures during pod-filling stages increase sucrose content. The raffinose contents of lines 8 and 9 in Group 5 (*RS2*/*RS3*/*mips1*) were notably lower than other genotypes across the 18 environments (Figures 4B and 4E). The environmental variation for raffinose content was smaller than sucrose and stachyose content. The reference check, line 10, in Group 6 (*RS2*/*RS3*/*MIPS1*) had the highest stachyose content among genotypes (Figures 4C and 4F). Notably, lines 8 and 9 in Group 5 (*RS2*/*RS3*/*mips1*) consistently showed lower stachyose contents than other genotypes across the 18 environments.

Germination and field emergence tests

The SG test showed a range of germination rates from 85.6% to 93.7% (Figure 5). R07-2000, a line with *RS2*/*RS3*/*mips1* allele combination, showed a significantly higher germination rate than the check, cv. Williams 82 (*RS2*/*RS3*/*MIPS1*), while the other lines tested were not significantly different than the check. The germination rates of a line with *RS2*/*rs3- Tilling*/*MIPS1* and the check were drastically dropped in the AA test, while lines with *rs2W331-*/*RS3*/*MIPS1* and *RS2*/*RS3*/*mips1* had high germination rates (87.5% for both). Field emergence tests in the BREC and FDREEC showed no significant difference among testing genotypes. The overall germination rates in the BREC were from 63.5% (*RS2*/*RS3*/*mips1*) to 74.5% (*RS2*/*rs3-Tilling*/*MIPS1*). The overall germination rates in the FDREEC drastically decreased and ranged from 18% (*RS2*/*RS3*/*mips1*) to 29.5%

(*RS2*/*RS3*/*MIPS1*). Notably, the line with the mips1 mutant allele showed a relatively lower germination rate than other testing lines in both field emergence tests.

DISCUSSION

The environmental stability and the genetic effect of individual and/or the combinations of mutant alleles, *RS2* and *RS3*, have been extensively investigated in soybean (Hagely et al., 2013; Bilyeu and Wiebold, 2016; Jo et al., 2018, 2019). However, *MIPS1* mutant soybean lines have not been tested to date. Therefore, this study benefits soybean breeders in understanding the environmental stability and genetic effect of the *mips1* allele. In addition, soybean breeders will have expanded germplasm sources to develop new soybean cultivars and germplasms with desirable carbohydrate profiles.

The results of our study confirmed the previous research on the correlation between soluble carbohydrate content and cooler temperature during pod-filling stages and the environmental effects, including year, location, and planting date on soluble carbohydrate content (Wolf et al., 1982; Bellaloui et al., 2010; Jaureguy et al., 2013; Bilyeu and Wiebold, 2016; Jo et al., 2018, 2019). Although the correlation was found in all genotypes tested, two soybean lines with *mips1* mutation showed more dramatic increases in sucrose content under cooler environments (Figure 1). The limitation of this study was that the planting dates and the days to maturity of genotypes were omitted. Late maturity lines could be exposed to lower temperatures for longer, possibly causing a higher sensitivity in sucrose content to cooler environments (Hou et al., 2009). Interestingly, significantly higher stachyose contents were observed in Portageville, MO, although the temperatures were intermediate among the testing locations during the pod-filling stages (Table 5). The trend of stachyose content varied in different research groups. Kumar et al. (2010) and Jo et al. (2018) reported stachyose contents were genotype-dependent, while Bilyeu and Wiebold (2016) found a positive correlation between temperature and stachyose content. Further

research will be required to enhance the current understanding of the relationship between carbohydrate content and environmental factors.

In the present study, soybean lines with *mips1* mutation had the most desirable seed carbohydrate profiles for animal feed, which was 11.1% sucrose, 0.53% raffinose, and 0.6% stachyose content (Figure 2). The other soybean lines with *rs2* and *rs3* mutations had 5.7% – 6.9% sucrose, 0.94% – 1.10% raffinose, and 1.6% – 2.6% stachyose content. However, as previously reported by Jo et al. (2018), the soybean line with *rs2W331-*/*rs3snp6*/*MIPS1* showed almost 10% sucrose and around 0.1% RFOs content. The phenotypic discrepancies between the same mutant lines in the present and previous studies were possibly derived from the different testing locations and years. Moreover, it was possible that such phenotypic discrepancy in soluble carbohydrate profiles would be the difference in HPLC equipment and the protocols. Nonetheless, the comparison for environmental stability and genetic effect within the current study should be scientifically meaningful because all genotypes were tested under the same circumstances across multiple locations and years using the same phenotyping protocol (Valliyodan et al., 2015).

The overall results in three germination tests showed no significant reduction in germination rates of the mutant lines carrying *rs2*, *rs3*, and *mips1* mutation, respectively, compared to the check cultivar, Williams 82. Interestingly, genotypes carrying *rs2W331-*/*RS3*/*MIPS1* and *RS2*/*RS3*/*mips1* maintained high germination rates in the AA test, while the other genotypes showed a significant reduction in germination rates. Although seeds used in the germination tests were obtained from a single environment at BREC in 2021, we would agree that further studies using multi-environment harvested seeds need to be conducted to test the effects of environments on the stability of germination. Moreover,

additional mutant genotypes carrying different allele combinations should be included to evaluate the effect of soluble carbohydrates on the aging of soybean seeds. A soybean line with *RS2*/*RS3*/*mips1* showed a relatively lower germination rate than other lines in the field emergence tests. This concern of poor field emergence among *mips1* mutant soybean lines was suggested by Meis et al. (2003) and Chappell et al. (2006). Since *MIPS1* gene is responsible for plant growth and development, phosphate storage, cell wall biosynthesis, cell-to-cell communication, and plant hormone transportation, the mutation in *MIPS1* gene may have negative impacts on soybean plants (Irvine and Schell, 2001). The current study compared the environmental stability and genetic effects of different mutant alleles conferring higher sucrose and lower RFOs. Although the lines with *mips1* mutation showed less environmental stability for sucrose content and lower field emergence rates in this study, field observations of R07-2000 (*mips1*) and/or R07-2001 (*mips1*) over the years in preliminary yield tests showed that seed germination and seedling establishment of these lines remained as good as the checks. In respect of value-added traits, these genotypes are suitable for animal and human uses due to the highest sucrose and lowest RFOs content; however, further breeding efforts are warranted to improve the environmental stability for sucrose and increase the field emergence rate of *mips1* mutant lines with the desirable carbohydrate profiles.

REFERENCES

Ajay, B.C., Aravind, J., and Abdul Fiyaz, R. (2019). Ammistability: R package for ranking genotypes based on stability parameters derived from AMMI model. *Indian Journal of Genetics and Plant Breeding*, 79(2), 460–466.

American Soybean Association. (2021). *Soystats*. Available at: http://soystats.com. Accessed 11 May 2023.

AOSA. (2002). Seed vigor testing handbook. *Contribution No. 32*. Association of Official Seed Analysts.

Bellaloui, N., Smith, J.R., Gillen, A.M., and Ray, J.D. (2010). Effect of maturity on seed sugars as measured on near-isogenic soybean (*Glycine max*) lines. *Crop Science*, 50(5), 1978–1987. https://doi.org/10.2135/cropsci2009.10.0596

Bilyeu, K., and 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(5), 1071–1078. https://doi.org/10.1021/acs.jafc.5b04779

Chappell, A.S., Scaboo, A.M., Wu, X., Nguyen, H., Pantalone, V.R., and Bilyeu, K.D. (2006). Characterization of the MIPS gene family in Glycine max. *Plant Breeding*, 125(5), 493–500. https://doi.org/10.1111/j.1439-0523.2006.01264.x

Coon, C.N., Leske, K.L., Akavanichan, O., and 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

Cunicelli, M.J., Bhandari, H.S., Chen, P., Sams, C.E., Rouf Mian, M.A., Mozzoni, L.A., et al. (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. http://doi.org/10.1002/aocs.12261

Dierking, E.C., and 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

Dierking, E.C., and 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

Fehr, W.R., Caviness, C.E., Burmood, D.T., and 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

Gauch, H.G. (1988). Model selection and validation for yield trials with interaction. *Biometrics*, 44(3), 705–715. https://doi.org/10.2307/2531585

Guillon, F., and Champ, M.M.-J. (2002). Carbohydrate fractions of legumes: Uses in human nutrition and potential for health. *British Journal of Nutrition*, 88, 293–306. https://doi.org/10.1079/BJN2002720

Hagely, K.B., Jo, H., Kim, J.H., Hudson, K.A., and 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., and Bilyeu, K. (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

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

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

Irvine, R.F., and Schell, M.J. (2001). Back in the water: The return of the inositol phosphates. *Nature Reviews Molecular Cell Biology*, 2, 327–338. https://doi.org/10.1038/35073015

Jaureguy, L.M., Rodriguez, F.L., Zhang, L., Chen, P., Brye, K., Oosterhuis, D., et al. (2013). Planting date and delayed harvest effects on soybean seed composition. *Crop Science*, 53(5), 2162–2175. https://doi.org/10.2135/cropsci2012.12.0683

Jo, H. (2016). Understanding sucrose and raffinose family of oligosaccharides in soybean seed for human and animal [Doctoral dissertation, University of Missouri]. University of Missouri.

Jo, H., Lee, J.-D., and 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, G., Chen, P., and 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

Karr-Lilienthal, L.K., Grieshop, C.M., Spears, J.K., and Fahey, G.C., Jr. (2005). Amino acid, carbohydrate, and fat composition of soybean meals prepared at 55 commercial U.S. soybean processing plants. *Journal of Agricultural and Food Chemistry*, 53(6), 2146–2150. https://doi.org/10.1021/jf048385i

Kerr, P.S., and Sebastian, S.A. (2000). Soybean products with improved carbohydrate composition and soybean plants. United States Patent 6,147,193.

Kumar, V., Rani, A., Goyal, L., Dixit, A.K., Manjaya, J.G., Dev, J., et al. (2010). Sucrose and raffinose family oligosaccharides (RFOs) in soybean seeds as influenced by

genotype and growing location. *Journal of Agricultural and Food Chemistry*, 58(8), 5081–5085. http://doi.org/10.1021/jf903141s

Le, H., Nguyen, N.H., Ta, D.T., Le, T.N.T., Bui, T.P., Le, N.T., et al. (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

Liu, K. (1997). Chemistry and nutritional value of soybean components. In K. Liu (Ed.), *Soybeans* (pp. 25-113). Springer US. https://doi.org/10.1007/978-1-4615-1763-4_2

Meis, S.J., Fehr, W.R., and Schnebly, S.R. (2003). Seed source effect on field emergence of soybean lines with reduced phytate and raffinose saccharides. *Crop Science*, 43(4), 1336–1339. https://doi.org/10.2135/cropsci2003.1336

Olivoto, T., and Lúcio, A.D. (2020). metan: An R package for multienvironment trial analysis. *Methods in Ecology and Evolution*, 11(6), 783–789. https://doi.org/10.1111/2041-210X.13384

R Core Team. (2021). R: A language and environment for statistical computing. *R Foundation for Statistical Computing*. Available at: https://www.Rproject.org/. Accessed 26 May 2023.

Saghai Maroof, M.A., and Buss, G.R. (2008). Low phytic acid, low stachyose, high sucrose soybean lines. U.S. Patent No. US8003856 B2.

Sebastian, S.A., Kerr, P.S., Pearlstein, R.W., and 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.

Singamsetti, A., Shahi, J.P., Zaidi, P.H., Seetharam, K., Vinayan, M.T., Kumar, M., et al. (2021). Genotype × environment interaction and selection of maize (*Zea mays* L.) hybrids across moisture regimes. *Field Crops Research*, 270, 108224. https://doi.org/10.1016/j.fcr.2021.108224

Skoneczka, J.A., Saghai Maroof, M.A., Shang, C., and 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

Valliyodan, B., Shi, H., and Nguyen, H.T. (2015). A simple analytical method for high-throughput screening of major sugars from soybean by normal-phase HPLC with evaporative light scattering detection. *Chromatography Research International*, 2015, 757649. https://doi.org/10.1155/2015/757649

Wolf, R.B., Cavins, J.F., Kleiman, R., and Black, L.T. (1982). Effect of temperature on soybean seed constituents: Oil, protein, fatty acids, amino acid, and sugar. *Journal of American Oil Chemists' Society*, 59(5), 230–232. https://doi.org/10.1007/BF02582182

TABLES AND FIGURES

Table 1. A list of soybean lines with pedigree, allelic status, and genotypic group.

Line	Name	Pedigree	Allelic status			Group[4]
			RS2[1]	RS3[2]	MIPS[3]	
1	CR13-395	KB07-15 × PI 542044	*rs2W331-*	*RS3*	*MIPS1*	1
2	CR13-397	KB07-15 × PI 542044	*rs2W331-*	*RS3*	*MIPS1*	
3	KB10-22#1548 b	Williams 82 EMS mutant × Patriot	*rs2T107I*	*RS3*	*MIPS1*	2
4	KB10-22#1568 b	Williams 82 EMS mutant × Patriot	*rs2T107I*	*RS3*	*MIPS1*	
5	KB10-22#1600 0	Williams 82 EMS mutant × Patriot	*rs2T107I*	*rs3snp6*	*MIPS1*	3
6	KB10-22#1608 0	Williams 82 EMS mutant × Patriot	*rs2T107I*	*rs3snp6*	*MIPS1*	
7	KB10-23#1681 0	KB07-15 × Patriot	*rs2W331-*	*rs3snp6*	*MIPS1*	4
8	R07-2000	Ozark × V99-5089	*RS2*	*RS3*	*mips1*	5
9	R07-2001	Ozark × V99-5089	*RS2*	*RS3*	*mips1*	
10	LD06-7620	IA3023 × LD00-3309	*RS2*	*RS3*	*MIPS1*	6

[1]*rs2W331-* and *rs2T107I* indicate mutant alleles for raffinose synthase 2 gene (*RS2*, *Glyma.06g179200*, Wm82.a2.v1); *RS2* indicates functional *RS2* which is similar to wild-type reference cultivar, LD06-7620. [2]*rs3snp6* indicates a mutant allele for raffinose synthase 3 gene (*RS3*, *Glyma.05g003900*, Wm82.a2.v1); *RS3* indicates functional *RS3* which is similar to wild-type reference cultivar, LD06-7620. [3]*mips1* indicates a mutant allele for D-myo-inositol-3-phosphate synthase 1 gene (*MIPS1*, *Glyma.11g238800*, Wm82.a2.v1); *MIPS1* indicates functional *MIPS1* which is similar to wild-type reference cultivar, LD06-7620. [4]Each genotypic group shares the same allele combination.

Table 2. A list of environment codes generated based on years, locations, and planting dates.

Environment code	Year	Location	Planting date[1]
E1	2015	Fayetteville, AR	Early
E2	2015	Fayetteville, AR	Late
E3	2015	Lafayette, IN	Early
E4	2015	Lafayette, IN	Late
E5	2015	Portageville, MO	Early
E6	2015	Portageville, MO	Late
E7	2016	Fayetteville, AR	Early
E8	2016	Fayetteville, AR	Late
E9	2016	Lafayette, IN	Early
E10	2016	Lafayette, IN	Late
E11	2016	Portageville, MO	Early
E12	2016	Portageville, MO	Late
E13	2017	Fayetteville, AR	Early
E14	2017	Fayetteville, AR	Late
E15	2017	Lafayette, IN	Early
E16	2017	Lafayette, IN	Late
E17	2017	Portageville, MO	Early
E18	2017	Portageville, MO	Late

[1]Planting date between 'Early' and 'Late' has an interval of about three weeks.

Table 3. Mean squares from analysis of variance for sucrose, raffinose, and stachyose compositions of 10 soybean lines grown in three locations at two planting dates over three years (2015-2017).

Source	df	Sucrose	Raffinose	Stachyose
Year (Y)	2	5.6***	0.54***	0.1
Location (L)	2	90.5***	0.44***	2.3***
Planting date (P)	1	5.8***	0.06	0.4
Y × L × P	8	1.1***	0.02	0.5***
Replication (Y × L × P)	18	0.6*	0.02	0.1
Genotype (G)	9	134.1***	3.46***	38.7***
G × Y	18	1.4***	0.06***	1.1***
G × L	18	1.5***	0.06***	1.8***
G × Y × L	36	1.2***	0.03**	0.4**
Error	243	0.3	0.02	0.2

*Significant at the $p < 0.05$ level. **Significant at the $p < 0.01$ level. ***Significant at the $p < 0.001$ level.

Table 4. Mean squares from analysis of variance for sucrose, raffinose, and stachyose compositions of six genotypic groups grown in three locations with two planting dates over three years (2015-2017).

Source	df	Sucrose	Raffinose	Stachyose
Year (Y)	2	4.9***	0.42***	0.1
Location (L)	2	70.9***	0.57***	2.3***
Planting date (P)	1	5.2**	0.06	0.5
Y × L × P	8	1.2*	0.02	0.6*
Replication (Y × L × P)	18	0.6	0.02	0.1
Genotypic group (Gg)	5	235.5***	6.16***	68.5***
Gg × Y	10	1.1*	0.07***	1.4***
Gg × L	10	2.4***	0.08***	2.8***
Gg × Y × L	20	1.1**	0.05***	0.4
Error	279	0.5	0.02	0.3

*Significant at the $p < 0.05$ level. **Significant at the $p < 0.01$ level. ***Significant at the $p < 0.001$ level.

Table 5. Phenotypic variation of sucrose, raffinose, and stachyose content among years, locations, and planting dates with monthly average temperatures during pod-filling stages.

Source	Sucrose	Raffinose	Stachyose	Temperature			
				July	August	September	October
	%			°C			
Year							
2015	7.3a	0.92a	1.9a	24.8	22.6	20.9	14.3
2016	6.8b	0.79b	2.0a	25.1	24.8	21.5	16.8
2017	7.3a	0.85ab	1.9a	25.2	22	20.3	14.8
Location							
Lafayette, IN	8.2a	0.79b	1.9b	22.3	21.3	19.5	13.4
Fayetteville, AR	6.5b	0.85ab	1.8b	26.6	24.4	22.1	16.4
Portageville, MO	6.7b	0.92a	2.1a	26.2	23.7	21.2	16.1
Planting date							
Early	7.0b	0.84a	2.0a			-	
Late	7.3a	0.86a	1.9a			-	

Note. Within a column, values followed by a different letter indicate significant differences at $p < 0.05$.

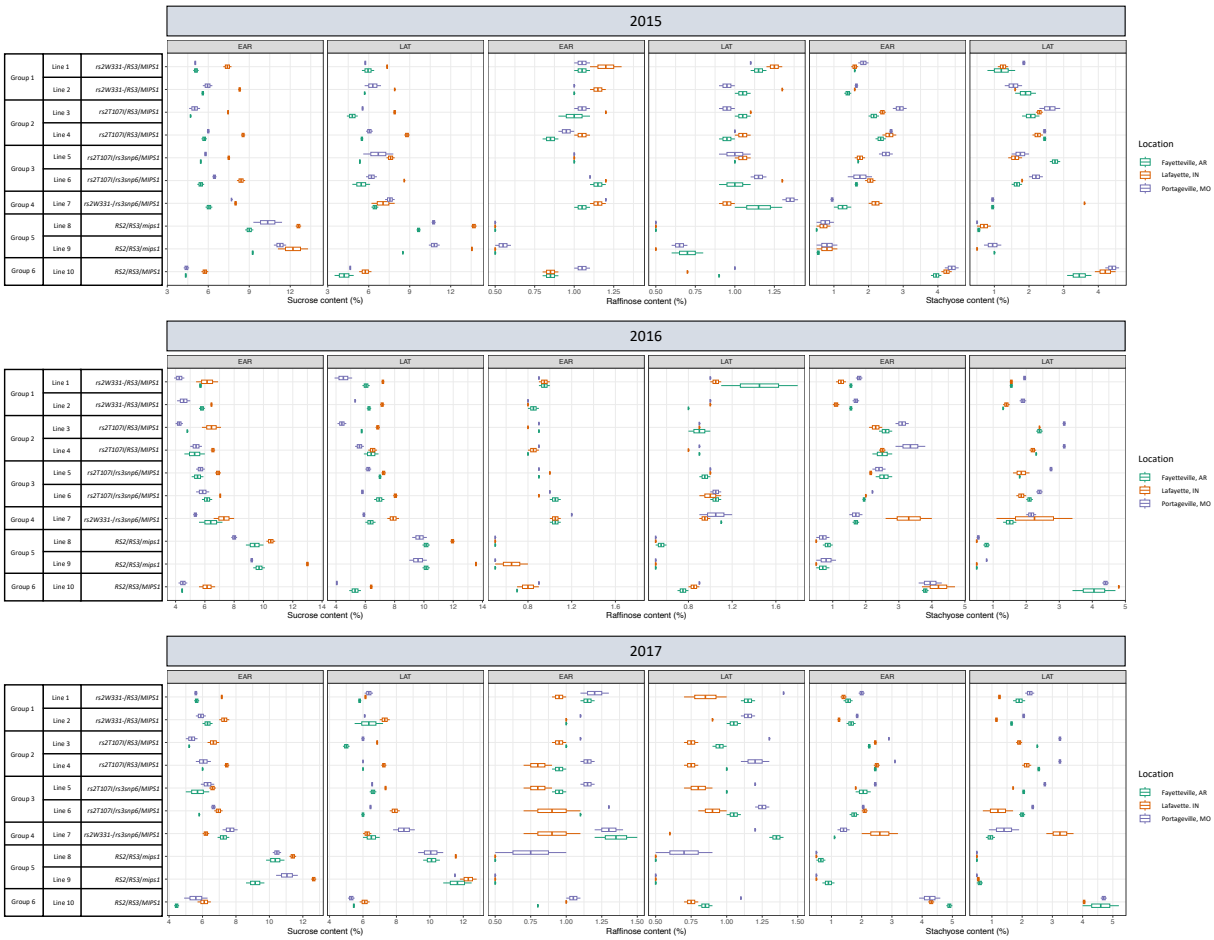


Figure 1. The mean values of sucrose, raffinose, and stachyose of ten genotypes in different years (2015 – 2017), locations (Fayetteville, Lafayette, and Portageville), and planting dates (early and late planting). EAR and LAT indicate early and late planting (three-week intervals), respectively.

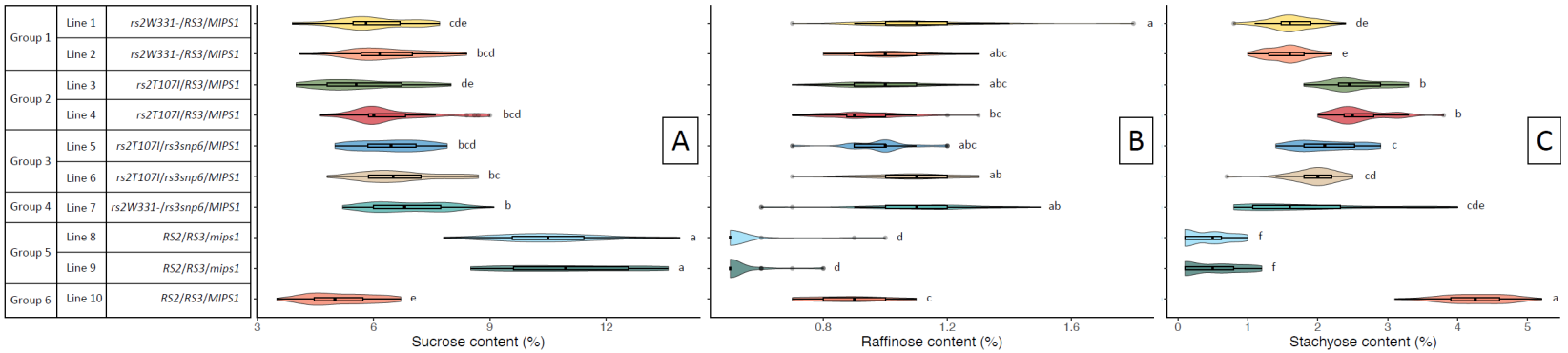


Figure 2. Sucrose, raffinose, and stachyose contents of ten soybean genotypes averaged across three locations and three years. Within each carbohydrate component, values not sharing a common letter are significantly different at $p < 0.05$.

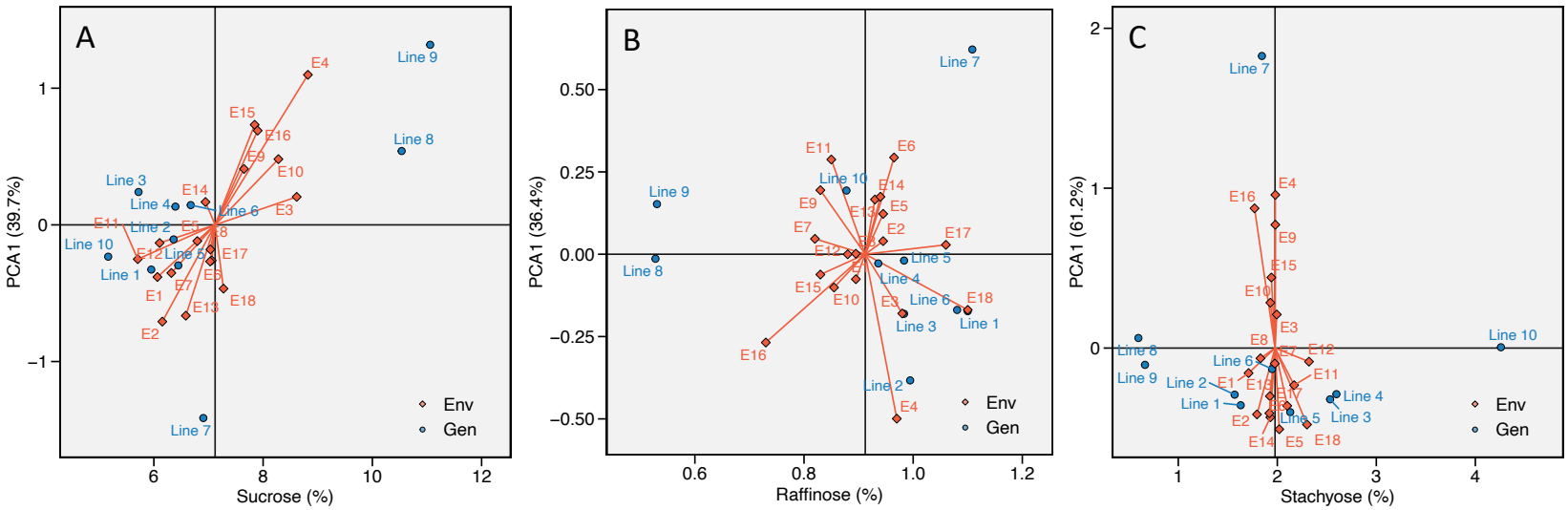


Figure 3. AMMI biplots for sucrose (A), raffinose (B), and stachyose (C) contents of ten genotypes across 18 environments. Lines 1 and 2 carry *rs2W331-/RS3/MIPS1* alleles, lines 3 and 4 carry *rs2T107I/RS3/MIPS1*, lines 5 and 6 carry *rs2T107I/rs3snp6/MIPS1*, a line 7 carries *rs2W331-/rs3snp6/MIPS1*, lines 8 and 9 carry *RS2/RS3/mips1*, and a line 10 carries *RS2/RS3/MIPS1*.

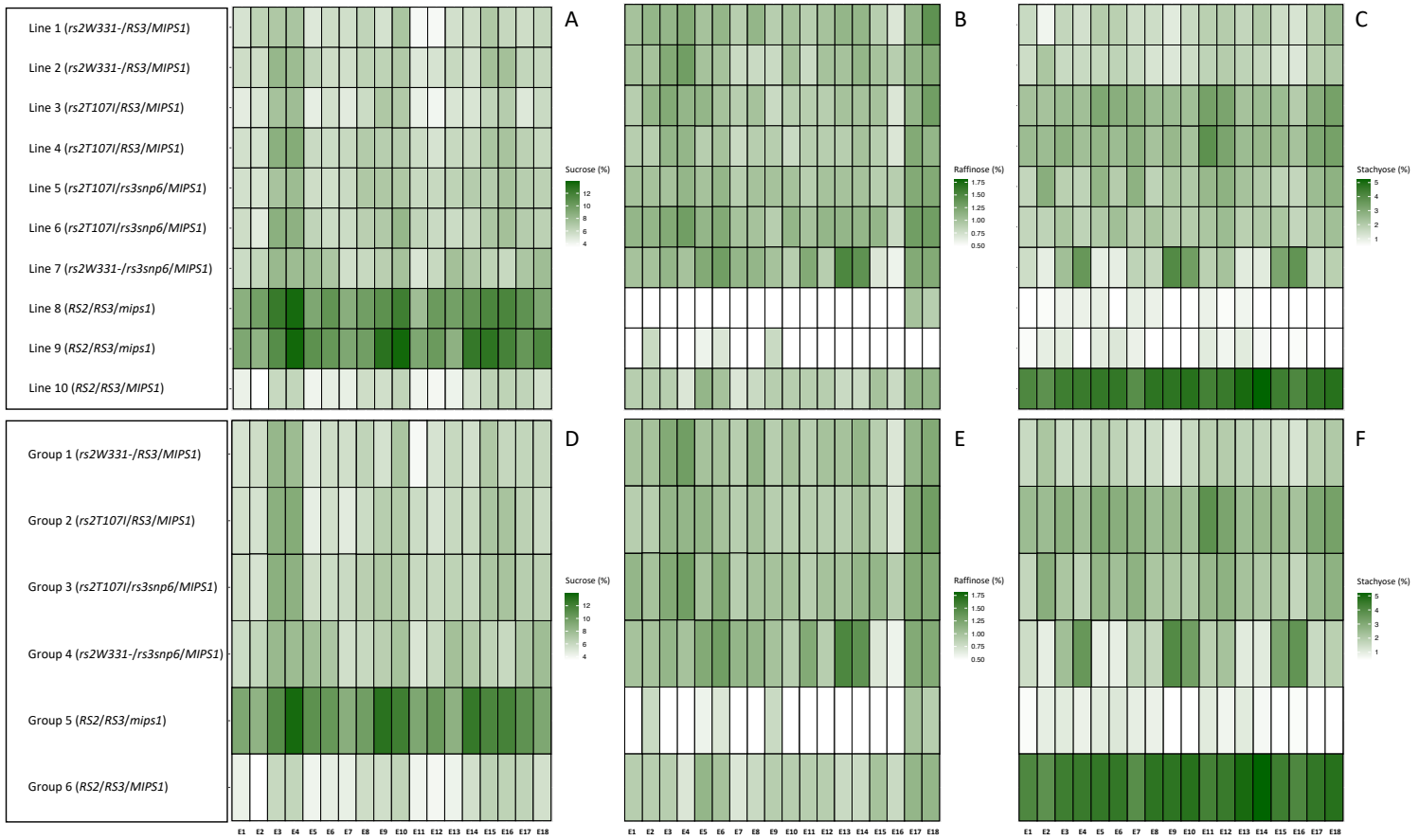


Figure 4. Heatmap for the sucrose (A), raffinose (B), and stachyose (C) content of ten genotypes and sucrose (D), raffinose (E), and stachyose (F) content of six genotypic groups across 18 individual environments.

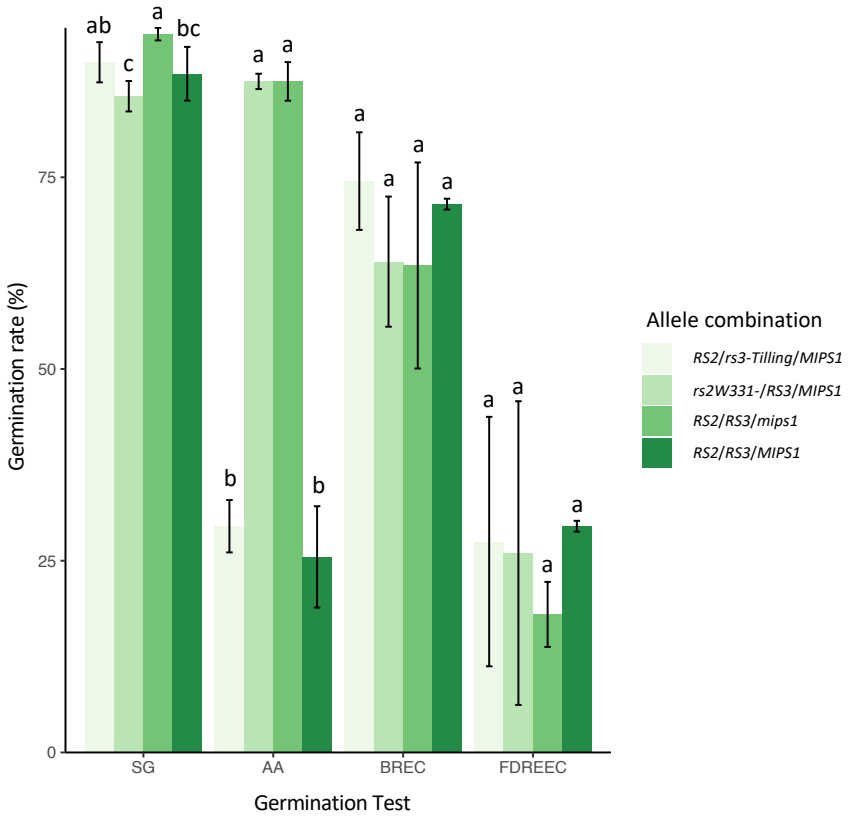


Figure 5. Germination rates (%) in standard germination (SG) test, accelerated aging (AA) test, and field emergence tests at two locations (BREC and FDREEC) of four genotypes, including 15045-13 (*RS2*/*rs3-Tilling*/*MIPS1*), KB07-15 (*rs2W331-*/*RS3*/*MIPS1*), R07-2000 (*RS2*/*RS3*/*mips1*), and Williams 82 (*RS2*/*RS3*/*MIPS1*). Within each test, values not sharing a common letter are significantly different at $p < 0.05$.

CONCLUSIONS AND FUTURE DIRECTIONS

Soluble carbohydrates in soybean are pivotal in animal feed efficiency and soy food products. Among soluble carbohydrates, sucrose is nutritionally beneficial, as it is easily digested and turned into metabolizable energy (ME). It also promotes the natural sweetness of soy foods, such as tofu, edamame, and soymilk. On the other hand, raffinose and stachyose (raffinose family of oligosaccharides, RFOs) are antinutrients that cause diarrhea and flatulence and reduce feed efficiency. This research had three main objectives: i) identifying novel genetic resources for high sucrose and low stachyose content in soybean through genome-wide association study (GWAS), ii) identifying novel variants for high sucrose content through quantitative trait loci (QTL) analysis and whole-genome sequencing (WGS) analysis, and iii) evaluating the genetic effect and environmental stability of known mutations conferring high sucrose and low RFOs in soybean.

In GWAS, we identified significant SNPs and candidate genes associated with sucrose and stachyose content in soybean. Given the diverse accessions, these genetic resources will be helpful to understand the genetic architecture underlying soluble carbohydrates in soybean. Also, we highlighted four candidate genes highly expressed in seed tissues during the seed development stages along with other plant parts, including seed coat, pod, leaf, and shoot. It suggested that those four genes might be closely associated with seed composition traits. Further investigations, including gene-editing techniques, will help validate the functions of candidate genes to increase sucrose and decrease stachyose content in soybean. Additionally, we reported promising soybean germplasm resources with desirable seed composition profiles for human and animal use.

This will give soybean breeders useful information about donor parents in breeding projects aiming at animal feed efficiency and soy food products.

With two bi-parental mapping populations sharing the same donor parent selected from the previous GWAS panel, we identified a major QTL region on Chr. 8, which confers high sucrose content in soybean. The subsequent WGS analysis for the major QTL region discovered novel SNP variants in three candidate genes closely related to transferring galactose to sucrose by constructing α-1,6-glycosidic bonds. The physical positions of SNP variants will be useful for molecular marker development, facilitating marker-assisted selection for high sucrose soybean lines in breeding programs. In addition, the donor parent, PI 506593, will be a promising germplasm in the breeding programs targeting the development of new food-grade soybean cultivars with high sucrose content and large seed size. For further investigation, high-throughput transcriptome sequencing (RNA-Seq) of our candidate genes will help us understand the gene expression patterns between parents during seed development. Also, the biological function of candidate genes can be further validated using genome editing technologies.

It is also important to test and utilize mutations previously identified for high sucrose and low RFOs content. Across 18 environments, lines carrying *mips1* mutation showed significantly higher sucrose and lower RFOs content than lines carrying *rs2* and *rs3* mutations. Despite a poor germination issue previously reported in *mips1* mutation, this study showed a line carrying *mip1* mutation had significantly greater germination rates than the check cultivar in the standard germination test and accelerated aging test and a slightly lower field emergence rate than the check. As *mips1* mutation was the most effective genetic resource to increase sucrose and reduce RFOs in soybean, further

breeding efforts will be required to sustain a desirable soluble carbohydrate profile and improve the field emergence rate. Also, this study will help soybean breeders choose high sucrose and low RFOs parental lines for breeding purposes.

APPENDIX

Supplementary tables and figures for Chapter 2

Supplementary Table 1. A list of accessions used for association study based on the USDA-ARS Soybean Germplasm Collection.

Line	Origin	Maturity group	100-seed weight
FC 19976-2	Japan	IV	38.0
PI 124871	Japan	IV	33.4
PI 157419	South Korea	IV	32.4
PI 157424	South Korea	IV	29.4
PI 196149	Japan	III	34.0
PI 196162	Japan	III	28.2
PI 19986	Japan	IV	28.6
PI 229343	Japan	IV	28.2
PI 243519	Japan	IV	35.4
PI 243527	Japan	IV	32.6
PI 243529	Japan	IV	31.2
PI 248514	Japan	IV	32.4
PI 274210	South Korea	IV	31.4
PI 339983	South Korea	IV	30.1
PI 339990	South Korea	IV	31.8
PI 342438	Japan	III	33.1
PI 398201	South Korea	IV	31.6
PI 398222	South Korea	IV	31.7
PI 398238	South Korea	V	30.4
PI 398240	South Korea	V	30.4
PI 398256	South Korea	IV	31.9
PI 398263	South Korea	V	32.9
PI 398293	South Korea	IV	32.9
PI 398319	South Korea	IV	30.8
PI 398342	South Korea	IV	30.3
PI 398379	South Korea	IV	28.9
PI 398401	South Korea	IV	32.0
PI 398450	South Korea	IV	31.4
PI 398531	South Korea	IV	31.2
PI 398532	South Korea	IV	31.8
PI 398615	South Korea	IV	32.5
PI 398735	South Korea	IV	31.1
PI 398738	South Korea	IV	29.7
PI 398744	South Korea	IV	29.7
PI 398759	South Korea	IV	29.6
PI 398767	South Korea	IV	29.0
PI 398779	South Korea	V	28.5
PI 398801	South Korea	IV	28.9
PI 398802	South Korea	IV	30.3
PI 398854	South Korea	IV	28.1
PI 398872	South Korea	IV	31.3
PI 398879	South Korea	IV	31.4
PI 398904	South Korea	IV	28.6
PI 398925	South Korea	VI	28.0
PI 399048	South Korea	VI	28.2
PI 399053	South Korea	VI	29.8
PI 399069	South Korea	IV	30.7
PI 407748	China	V	28.3
PI 408033	South Korea	IV	30.9
PI 408058	South Korea	IV	31.4

PI 408064	South Korea	IV	32.0
PI 408065	South Korea	IV	31.5
PI 408072	South Korea	IV	30.2
PI 408076B	South Korea	IV	34.2
PI 408109A	South Korea	IV	29.5
PI 408125B	South Korea	IV	31.1
PI 408179	South Korea	IV	29.3
PI 408228B	South Korea	IV	31.8
PI 408233B	South Korea	IV	29.9
PI 408254	South Korea	VI	30.7
PI 408263	South Korea	IV	32.0
PI 408291	South Korea	IV	30.4
PI 408298A	South Korea	IV	30.1
PI 408299	South Korea	IV	31.2
PI 408334	South Korea	IV	31.8
PI 416888	Japan	IV	33.9
PI 416892	Japan	III	37.3
PI 417006	Japan	IV	32.2
PI 417021	Japan	IV	31.6
PI 417086B	Japan	IV	30.5
PI 417099	Japan	VI	28.1
PI 417159	Japan	V	29.0
PI 417163	Japan	IV	34.4
PI 417233	Japan	IV	29.3
PI 417301	Japan	IV	30.4
PI 417322	Japan	V	29.3
PI 417339	Japan	IV	34.4
PI 417468	Japan	IV	34.4
PI 417491	Japan	V	29.1
PI 423739	South Korea	IV	28.8
PI 423740	South Korea	IV	29.1
PI 423743B	South Korea	V	29.0
PI 423750	South Korea	IV	29.4
PI 423763	South Korea	IV	31.0
PI 423777	South Korea	IV	29.1
PI 423830B	South Korea	IV	28.4
PI 423899	Japan	III	32.5
PI 423980	Japan	IV	32.4
PI 424139	South Korea	VI	30.0
PI 424151	South Korea	IV	29.5
PI 424185	South Korea	VI	32.1
PI 424189	South Korea	IV	29.8
PI 424233	South Korea	IV	33.3
PI 424237A	South Korea	IV	32.8
PI 424250A	South Korea	IV	28.1
PI 424255B	South Korea	IV	28.0
PI 424282	South Korea	IV	31.7
PI 424292	South Korea	IV	32.3
PI 424313	South Korea	IV	31.3
PI 424314	South Korea	IV	30.2
PI 424322	South Korea	IV	29.0
PI 424364A	South Korea	IV	31.0
PI 424364B	South Korea	IV	29.7
PI 424377	South Korea	IV	28.3
PI 424379	South Korea	IV	29.2
PI 424385	South Korea	IV	29.9

PI 424439	South Korea	VI	28.3
PI 424450	South Korea	IV	30.0
PI 424459	South Korea	IV	35.1
PI 424470	South Korea	IV	28.5
PI 424484B	South Korea	IV	29.7
PI 424492	South Korea	IV	31.0
PI 424493A	South Korea	IV	29.0
PI 424513	South Korea	IV	29.6
PI 424516	South Korea	IV	28.6
PI 424518	South Korea	IV	28.3
PI 424537	South Korea	IV	33.4
PI 424546A	South Korea	IV	32.2
PI 424558A	South Korea	IV	30.2
PI 424564	South Korea	IV	29.9
PI 424571	South Korea	IV	29.3
PI 424574	South Korea	IV	30.5
PI 424576	South Korea	V	29.6
PI 424586	South Korea	V	28.6
PI 424590A	South Korea	IV	31.0
PI 424590B	South Korea	IV	30.6
PI 437734	China	V	28.6
PI 438300	North Korea	IV	28.2
PI 442007B	South Korea	IV	28.4
PI 458021	South Korea	IV	31.1
PI 458031	South Korea	IV	29.8
PI 458036	South Korea	IV	28.6
PI 458055	South Korea	IV	29.7
PI 458086	South Korea	IV	28.1
PI 458118	South Korea	IV	29.1
PI 458125	South Korea	IV	29.2
PI 458136	South Korea	IV	28.2
PI 458141	South Korea	IV	31.2
PI 458159	South Korea	V	29.3
PI 458203A	South Korea	IV	31.8
PI 458203B	South Korea	IV	29.7
PI 458219	South Korea	VI	29.7
PI 458245	South Korea	VI	29.2
PI 504508	Japan	III	28.9
PI 506530	Japan	VI	30.0
PI 506560	Japan	IV	33.2
PI 506569	Japan	VI	30.7
PI 506592	Japan	III	37.8
PI 506593	Japan	VI	32.0
PI 506594	Japan	V	32.2
PI 506606	Japan	VI	29.0
PI 506637	Japan	III	36.5
PI 506730	Japan	V	30.6
PI 506746	Japan	V	29.3
PI 506752	Japan	V	28.4
PI 506789	Japan	IV	29.5
PI 506790	Japan	III	32.5
PI 506797	Japan	V	31.2
PI 506799	Japan	III	33.4
PI 506800A	Japan	III	34.3
PI 506800B	Japan	III	34.0
PI 506801A	Japan	III	29.8

PI 506801B	Japan	III	32.7
PI 506890	Japan	V	28.2
PI 506903	Japan	IV	36.1
PI 506937	Japan	IV	29.0
PI 506982	Japan	III	29.7
PI 506987	Japan	III	28.0
PI 506993	Japan	IV	31.7
PI 507031	Japan	V	29.1
PI 507121	Japan	V	31.0
PI 507123	Japan	IV	28.8
PI 507179	Japan	IV	30.4
PI 507208	Japan	VI	29.0
PI 507226A	Japan	III	35.3
PI 507226B	Japan	III	33.6
PI 507273	Japan	III	29.3
PI 507309	Japan	IV	33.6
PI 507428	Japan	VI	28.2
PI 507433	Japan	V	28.4
PI 507438	Japan	VI	29.9
PI 507445	Japan	IV	31.3
PI 507449	Japan	IV	29.3
PI 507469	Japan	VI	32.7
PI 507487	Japan	III	35.7
PI 507523	Japan	III	32.2
PI 507570	Japan	III	33.5
PI 509086	South Korea	VI	28.2
PI 509093	South Korea	VI	30.8
PI 536547B	Taiwan	III	30.6
PI 536547C	Taiwan	III	29.8
PI 548350	Japan	IV	31.8
PI 548351	South Korea	IV	32.8
PI 548361	Japan	III	31.0
PI 548408	China	IV	28.1
PI 548457	Japan	VI	28.2
PI 548486	Japan	VI	36.9
PI 548559	US	IV	31.3
PI 548587	US	III	28.6
PI 548624	US	III	34.6
PI 549063	Japan	IV	28.4
PI 549065	Japan	IV	34.5
PI 561288	Taiwan	IV	32.4
PI 561292A	Taiwan	III	33.5
PI 561292B	Taiwan	IV	31.8
PI 561293	Taiwan	IV	31.6
PI 567764	China	V	28.9
PI 578467	China	VI	30.6
PI 70243	China	IV	30.6
PI 80459	Japan	III	28.8
PI 81029-1	Japan	IV	29.7
PI 85441	South Korea	IV	32.0
PI 86134-1	Japan	IV	30.8
PI 86134-4	Japan	IV	28.3
PI 87165	US	III	34.4
PI 89162	North Korea	III	28.5
PI 91684	North Korea	IV	30.5
PI 96118	North Korea	IV	28.0

PI 96550	North Korea	IV	28.2
PI 96783	North Korea	IV	28.9

Supplementary Table 2. Correlations between sucrose and stachyose content in seven environments

Environment	Correlation
FAY_14	-0.11*
STU_14	0.33***
FAY_15	-0.1*
STU_15	-0.12*
POR_20	0.01^{ns}
COL_21	-0.08^{ns}
POR_21	-0.05^{ns}

^{ns}not significant. *significant at $p < 0.05$. **significant at $p < 0.01$. ***significant at $p < 0.001$.

Supplementary Table 3. Mean squares from analysis of variance for the sucrose and stachyose content.

Source	DF	Sucrose	Stachyose
Environment (E)	6	100.2***	53.3***
Maturity Group (M)	3	38.8***	190.5***
Genotype within M (G)	216	5.7***	4.1***
G × E	1314	0.6***	0.3***
Replication within Genotype	220	0.1^{ns}	0.2^{ns}

^{ns}not significant. *significant at $p < 0.05$. **significant at $p < 0.01$. ***significant at $p < 0.001$.

Supplementary Table 4. A list of SNPs identified through genome-wide association study across all environments.

Trait	SNP	Chromosome	Position (bp)	-log(P)	Environment[1]
Sucrose	ss715581176	2	13472284	3.8	STU_14
	ss715581176	2	13472284	3.6	STU_15
	ss715581177	2	13478261	3.9	STU_14
	ss715581177	2	13478261	3.9	STU_15
	ss715581177	2	13478261	3.0	BLUP
	ss715581178	2	13479724	3.8	STU_14
	ss715581178	2	13479724	3.6	STU_15
	ss715581183	2	13523639	4.0	STU_14
	ss715581183	2	13523639	3.9	STU_15
	ss715581183	2	13523639	3.6	BLUP
	ss715581345	2	15473746	3.0	POR_20
	ss715582485	2	39808726	3.1	FAY_15
	ss715582936	2	42935784	3.2	POR_21
	ss715584684	3	1996434	3.4	FAY_15
	ss715585782	3	3929449	3.2	STU_14
	ss715585782	3	3929449	3.3	BLUP
	ss715586772	3	5081945	3.3	COL_21
	ss715588304	4	469489	3.0	FAY_14

ss715591544	5	5486958	3.1	COL_21
ss715591544	5	5486958	3.3	POR_21
ss715592443	5	2372200	3.0	BLUP
ss715592444	5	2374822	3.1	FAY_14
ss715598115	7	42459910	3.4	COL_21
ss715598115	7	42459910	3.0	BLUP
ss715598116	7	42477713	3.1	COL_21
ss715599696	8	15317735	3.4	STU_15
ss715599738	8	15769392	3.1	POR_20
ss715599828	8	16525923	3.0	STU_15
ss715601987	8	42465843	3.2	STU_14
ss715601989	8	42485714	3.2	STU_14
ss715602502	8	47286262	3.7	FAY_14
ss715602502	8	47286262	3.8	FAY_15
ss715602502	8	47286262	3.1	STU_15
ss715602502	8	47286262	4.1	COL_21
ss715602502	8	47286262	4.8	BLUP
ss715602503	8	47317763	3.4	FAY_14
ss715602503	8	47317763	3.7	COL_21
ss715602503	8	47317763	3.9	BLUP
ss715602505	8	47332629	3.7	COL_21
ss715602505	8	47332629	3.2	POR_21
ss715604115	9	41717894	3.3	STU_14
ss715604752	9	47006196	3.0	COL_21
ss715604753	9	47007331	3.6	COL_21
ss715604753	9	47007331	3.2	BLUP
ss715608144	10	50562912	3.4	POR_20
ss715608145	10	50564648	3.3	POR_20
ss715612395	12	34250868	3.1	STU_15
ss715612415	12	34382741	3.1	STU_15
ss715612433	12	34489617	3.0	STU_15
ss715613179	12	5486355	3.9	FAY_15
ss715613179	12	5486355	4.0	COL_21
ss715613179	12	5486355	3.2	POR_21
ss715613179	12	5486355	3.9	BLUP
ss715613311	12	6694566	3.4	POR_21
ss715613553	12	828691	3.1	FAY_14
ss715613559	12	87780	3.5	COL_21
ss715613588	12	855537	3.4	FAY_14
ss715613588	12	855537	3.1	BLUP
ss715613589	12	856720	3.1	FAY_14
ss715613589	12	856720	3.3	BLUP
ss715613592	12	858009	3.0	BLUP
ss715613645	12	94121	3.3	COL_21
ss715613685	12	949609	3.1	FAY_14
ss715613685	12	949609	3.0	STU_15
ss715613685	12	949609	3.3	BLUP
ss715614762	13	29177321	3.3	STU_15
ss715616020	13	18159747	3.1	FAY_14
ss715616989	13	15130627	3.1	COL_21
ss715617844	14	1672754	4.2	POR_21
ss715617894	14	1795304	3.5	POR_21
ss715619786	14	6693932	3.0	STU_15
ss715619831	14	7089189	3.0	STU_15
ss715624558	16	32840492	3.2	FAY_15
ss715624558	16	32840492	3.3	POR_20
ss715625545	16	778692	3.1	POR_20
ss715625552	16	781158	3.0	POR_20
ss715625556	16	782140	3.1	POR_20
ss715626916	17	36094194	3.4	FAY_14
ss715626940	17	36165859	3.7	FAY_14
ss715627190	17	37039145	3.8	FAY_14

	ss715627314	17	37876098	3.3	FAY_14
	ss715627317	17	37877925	3.1	FAY_14
	ss715627820	17	41098767	4.2	FAY_14
	ss715627820	17	41098767	3.7	STU_14
	ss715627820	17	41098767	3.4	BLUP
	ss715627837	17	41319215	3.3	POR_21
	ss715627851	17	41433652	3.3	FAY_15
	ss715627851	17	41433652	4.2	POR_21
	ss715627851	17	41433652	3.4	BLUP
	ss715627853	17	41440620	3.5	FAY_14
	ss715627853	17	41440620	4.1	COL_21
	ss715627853	17	41440620	5.2	POR_21
	ss715627853	17	41440620	4.3	BLUP
	ss715629735	18	2036179	3.2	STU_15
	ss715630834	18	43812070	3.4	STU_14
	ss715631854	18	53019579	3.1	POR_21
	ss715632574	18	6908475	3.7	FAY_14
	ss715632574	18	6908475	3.3	BLUP
	ss715632576	18	6923834	3.3	FAY_14
	ss715632579	18	6946338	3.2	FAY_14
	ss715632579	18	6946338	3.3	BLUP
	ss715632580	18	6948950	3.0	FAY_14
	ss715632584	18	7012979	3.2	BLUP
	ss715633619	19	2901793	3.3	COL_21
	ss715633626	19	2916347	3.2	COL_21
	ss715633626	19	2916347	3.0	POR_21
	ss715633649	19	2978526	3.0	COL_21
	ss715633665	19	3022199	3.3	COL_21
	ss715633671	19	3030986	3.1	COL_21
	ss715635124	19	41536524	3.1	STU_15
	ss715635645	19	47208164	3.0	POR_20
	ss715636768	20	1655457	3.0	POR_21
	ss715636770	20	1657987	3.2	POR_21
	ss715636771	20	1658893	3.3	POR_21
	ss715636780	20	1672879	3.1	POR_21
	ss715636781	20	1678458	3.1	POR_21
	ss715636857	20	1907881	3.7	COL_21
	ss715636857	20	1907881	3.1	POR_21
	ss715636857	20	1907881	3.1	BLUP
	ss715636867	20	1937812	3.3	COL_21
	ss715636869	20	1940193	3.5	COL_21
	ss715637424	20	34276503	3.1	FAY_15
	ss715637427	20	34285042	3.1	FAY_15
	ss715637428	20	34286637	3.4	FAY_15
	ss715637428	20	34286637	3.3	STU_15
	ss715637428	20	34286637	3.6	BLUP
	ss715638979	20	541850	3.3	COL_21
	ss715638979	20	541850	3.2	POR_21
	ss715638979	20	541850	3.1	BLUP
Stachyose	ss715579934	1	4934180	3.0	STU_15
	ss715579872	1	49445573	3.0	STU_15
	ss715579880	1	49497388	3.0	STU_15
	ss715580039	1	50700924	3.0	POR_21
	ss715582428	2	39567015	3.2	POR_21
	ss715583079	2	44214908	3.7	STU_15
	ss715583079	2	44214908	3.5	POR_20
	ss715583079	2	44214908	3.3	BLUP
	ss715583083	2	44223082	3.2	STU_15
	ss715583117	2	44445656	3.1	POR_20
	ss715583119	2	44448179	3.1	FAY_14
	ss715583119	2	44448179	3.8	FAY_15
	ss715583119	2	44448179	3.9	STU_15

ss715583119	2	44448179	3.2	POR_20
ss715583119	2	44448179	3.1	COL_21
ss715583119	2	44448179	3.9	BLUP
ss715583121	2	44464449	3.1	POR_20
ss715583124	2	44478874	3.1	POR_20
ss715583125	2	44496502	3.1	POR_20
ss715583127	2	44505896	3.7	STU_15
ss715583127	2	44505896	3.4	POR_20
ss715583127	2	44505896	3.3	BLUP
ss715583129	2	44546704	3.2	STU_14
ss715583129	2	44546704	3.0	FAY_15
ss715583129	2	44546704	3.3	STU_15
ss715583129	2	44546704	3.4	POR_20
ss715583129	2	44546704	3.1	BLUP
ss715583131	2	44553129	3.2	FAY_15
ss715583131	2	44553129	3.4	STU_15
ss715583131	2	44553129	3.5	POR_20
ss715583131	2	44553129	3.5	BLUP
ss715586461	3	4411955	3.0	POR_21
ss715585362	3	33694405	3.3	STU_15
ss715585426	3	34244402	3.0	COL_21
ss715585880	3	38086799	3.3	STU_14
ss715588402	4	46974476	3.7	FAY_14
ss715588402	4	46974476	3.4	FAY_15
ss715588405	4	46994257	3.3	FAY_15
ss715588530	4	47858859	3.3	COL_21
ss715588554	4	48133794	3.1	COL_21
ss715588625	4	48507459	3.1	STU_14
ss715588631	4	48530066	3.0	COL_21
ss715592317	5	2167334	3.7	STU_14
ss715592317	5	2167334	3.6	COL_21
ss715592340	5	2207089	3.4	STU_14
ss715592340	5	2207089	3.4	COL_21
ss715592340	5	2207089	3.0	BLUP
ss715592396	5	2292131	3.3	STU_14
ss715592436	5	2362636	3.9	STU_14
ss715592436	5	2362636	3.1	FAY_15
ss715592437	5	2363027	3.5	STU_14
ss715592438	5	2363390	3.5	FAY_14
ss715592439	5	2363598	4.5	STU_14
ss715592439	5	2363598	3.1	COL_21
ss715592439	5	2363598	3.5	BLUP
ss715592440	5	2365275	4.0	STU_14
ss715592440	5	2365275	3.3	FAY_15
ss715592440	5	2365275	3.5	BLUP
ss715592442	5	2369980	4.2	FAY_14
ss715592442	5	2369980	4.9	STU_14
ss715592442	5	2369980	3.8	FAY_15
ss715592442	5	2369980	3.2	STU_15
ss715592442	5	2369980	3.4	POR_20
ss715592442	5	2369980	3.4	COL_21
ss715592442	5	2369980	5.0	BLUP
ss715589992	5	3263938	3.1	POR_20
ss715590248	5	4269814	3.2	POR_21
ss715591198	5	35773064	3.0	FAY_14
ss715591198	5	35773064	3.1	FAY_15
ss715591198	5	35773064	3.4	POR_20
ss715591198	5	35773064	3.3	BLUP
ss715593030	6	13368203	3.7	FAY_14
ss715593403	6	15754774	3.5	STU_15
ss715593403	6	15754774	3.5	BLUP
ss715594464	6	44083402	3.2	FAY_15

ss715596074	7	11585878	3.0	FAY_14
ss715596614	7	17022485	3.1	FAY_14
ss715601133	8	2382910	3.1	STU_15
ss715601133	8	2382910	3.4	POR_20
ss715601133	8	2382910	3.4	POR_21
ss715601133	8	2382910	3.4	BLUP
ss715601165	8	2470886	3.1	COL_21
ss715601341	8	2983923	3.2	COL_21
ss715603666	9	3435794	3.1	COL_21
ss715603671	9	3446653	3.5	FAY_14
ss715603873	9	3766363	3.1	FAY_14
ss715603880	9	3771212	4.0	FAY_14
ss715603880	9	3771212	3.0	STU_14
ss715603880	9	3771212	3.5	BLUP
ss715603882	9	3772351	3.4	FAY_14
ss715603882	9	3772351	3.1	BLUP
ss715605014	9	4680450	3.3	COL_21
ss715605022	9	4693630	3.1	COL_21
ss715605024	9	4695047	3.6	COL_21
ss715605025	9	4696895	3.1	COL_21
ss715605035	9	4706268	3.0	COL_21
ss715605079	9	4874816	3.6	COL_21
ss715605274	9	6311012	3.2	FAY_15
ss715605274	9	6311012	3.3	POR_20
ss715603435	9	29576509	3.2	POR_21
ss715639178	9	30134957	3.0	FAY_14
ss715639178	9	30134957	3.3	COL_21
ss715639178	9	30134957	3.8	POR_21
ss715639178	9	30134957	3.3	BLUP
ss715606330	10	3541231	3.0	FAY_14
ss715606330	10	3541231	3.6	STU_14
ss715606330	10	3541231	3.4	FAY_15
ss715615801	13	18315025	3.1	FAY_14
ss715615801	13	18315025	3.3	STU_15
ss715615801	13	18315025	3.3	POR_20
ss715615801	13	18315025	3.9	POR_21
ss715615801	13	18315025	3.6	BLUP
ss715615716	13	18379941	3.1	FAY_14
ss715615716	13	18379941	3.3	STU_15
ss715615716	13	18379941	3.3	POR_20
ss715615716	13	18379941	3.9	POR_21
ss715615716	13	18379941	3.6	BLUP
ss715614101	13	24090619	3.3	FAY_15
ss715614101	13	24090619	3.0	COL_21
ss715614101	13	24090619	3.1	BLUP
ss715615379	13	33065973	3.2	POR_21
ss715617647	14	12948873	3.3	COL_21
ss715617652	14	13008001	3.2	COL_21
ss715617670	14	13134423	3.1	COL_21
ss715617671	14	13139864	3.5	COL_21
ss715617672	14	13154645	3.0	POR_20
ss715617672	14	13154645	3.7	COL_21
ss715617672	14	13154645	3.2	BLUP
ss715617675	14	13187218	3.3	POR_20
ss715617675	14	13187218	3.9	COL_21
ss715617675	14	13187218	3.5	BLUP
ss715617677	14	13213128	3.8	COL_21
ss715617678	14	13234472	4.1	COL_21
ss715617679	14	13277133	3.7	COL_21
ss715617681	14	13297944	3.3	COL_21
ss715617682	14	13304518	3.2	COL_21
ss715617684	14	13312001	3.2	COL_21

ss715617686	14	13319757	3.5	COL_21
ss715617687	14	13321438	3.4	COL_21
ss715617690	14	13330903	3.7	COL_21
ss715617691	14	13332047	3.2	COL_21
ss715617697	14	13358759	3.2	COL_21
ss715622804	15	51605937	3.1	FAY_15
ss715622806	15	51630810	3.2	FAY_15
ss715622806	15	51630810	3.0	POR_20
ss715622806	15	51630810	3.1	BLUP
ss715623995	16	28898796	3.1	STU_14
ss715624036	16	29107753	3.0	STU_14
ss715626641	17	28887667	3.0	STU_15
ss715626867	17	35745764	3.3	BLUP
ss715631422	18	48872732	3.2	POR_20
ss715634825	19	39500216	3.2	STU_15
ss715634825	19	39500216	3.3	BLUP

[1]FAY_14 (Fayetteville in 2014), STU_14 (Stuttgart in 2014), FAY_15 (Fayetteville in 2015), STU_15 (Stuttgart in 2015), POR_20 (Portageville in 2020), COL_21 (Columbia in 2021), and POR_21 (Portageville in 2021)

Supplementary Table 5. A list of tentative candidate genes located within the LD block of significant SNPs.

Trait	Significant SNP	Chr	LD	Candidate gene[1]	Function Annotation[2]	Protein interaction[3]
Sucrose	ss715581183	2	13130661..13523639	*Glyma.02g128500*	Predicted membrane protein (DUF2232)	-[4]
				Glyma.02g128800	Protein of unknown function (DUF1262)	F-box/kelch-repeat protein at1g23390
				Glyma.02g128900	Importin-beta N-terminal domain	Nucleosome assembly protein 1
				Glyma.02g129000	Importin - 7, 8, 11	Nucleosome assembly protein 1
				Glyma.02g129100	Helix-loop-helix DNA-binding domain	Retinoblastoma-related protein 1
				Glyma.02g129200	Predicted hydrolases of HD superfamily	Xyloglucan galactosyltransferase
				Glyma.02g129300	Leucine-rich repeat receptor-like protein kinase	E3 ubiquitin-protein ligase
				Glyma.02g129500	Haloacid dehalogenase-like hydrolase	Ala-interacting subunit 3 isoform x1
				Glyma.02g129700	Protein of unknown function	F-box/kelch-repeat protein at1g23390
				Glyma.02g129800	Cytidine and deoxycytidylate deaminase zinc-binding region	Trna-specific adenosine deaminase tad3 isoform x1
				Glyma.02g130100	DNL zinc finger	Stromal 70 kda heat shock-related protein
				Glyma.02g130200	Xenotropic and polytropic retrovirus receptor 1	Kinesin-like protein kin-12c isoform x1
				Glyma.02g130400	Flavonoid biosynthesis	Cationic peroxidase 1
				Glyma.02g130600	ATP-binding cassette transporter	Glycerol-3-phosphate acyltransferase
				Glyma.02g130700	CO_2 fixation into oxaloacetate	Methylcrotonoyl-coa carboxylase beta chain, mitochondrial isoform x1
				Glyma.02g131000	Endonuclease/Exonuclease/phosphatase family	Methylcrotonoyl-coa carboxylase beta chain, mitochondrial isoform x1
	ss715602502	8	46921284..47286262	*Glyma.08g356300*	Protein of unknown function	Exocyst complex component exo70e2
				Glyma.08g356500	Serine-Threonine protein kinase	Serine/threonine-protein kinase mtor; Belongs to the PI3/PI4-kinase family
				Glyma.08g356600	Early growth response protein-related	-
				Glyma.08g356700	Tetratricopeptide repeat (TPR)-like superfamily protein	Serine/threonine-protein kinase edr1; Uncharacterized protein
				Glyma.08g356800	Pectin lyase-like superfamily protein	Polygalacturonase
				Glyma.08g356900	Helix-loop-helix DNA-binding domain	E3 ubiquitin-protein ligase
				Glyma.08g357000	ER to golgi transport protein/RAD50-interacting protein 1	Centromere/kinetochore protein zw10 homolog isoform x1
				Glyma.08g357100	Acetylglucosaminyltransferase family protein	Lipid transfer-like protein vas
				Glyma.08g357200	Serine-Threonine protein kinase	1D-myo-inositol 3-kinase
				Glyma.08g357400	Serine-Threonine protein kinase	Leucine-rich repeat receptor protein kinase ems1
				Glyma.08g357500	Protein tyrosine kinase	Kinase-associated protein phosphatase 2
				Glyma.08g357600	AP2/B3-like transcriptional factor family protein	E3 ubiquitin-protein ligase
				Glyma.08g358200	ARM repeat superfamily protein	Proteasome subunit beta type
				Glyma.08g358300	26S proteasome regulatory complex, subunit PSMD10	Protein phosphatase 2c and cyclic nucleotide-binding/kinase domain-containing protein isoform x1

			Glyma.08g358700	WD40 repeat-containing protein	Mannosyl-oligosaccharide alpha-1,3-glucosidase / glycosyl hydrolase 31 family
			Glyma.08g358800	D-galactose detoxification	UTP--glucose-1-phosphate uridylyltransferase / UDP-sugar pyrophosphorylase
			Glyma.08g359000	Embryo-specific protein 3	10 kDa heat shock protein
			Glyma.08g359100	Multicopper oxidase	Plastocyanin
			Glyma.08g359200	Multicopper oxidase	Plastocyanin
			Glyma.08g359300	Multicopper oxidase	Plastocyanin
			Glyma.08g359400	Multicopper oxidase	Plastocyanin
			Glyma.08g359700	CemA-like proton extrusion protein-related	Pentatricopeptide repeat-containing protein at3g46790
			Glyma.08g359900	Activator of Hsp90 ATPase, N-terminal	Hsp70 nucleotide exchange factor fes1
			Glyma.08g360100	LITAF-like zinc ribbon domain	Transcription factor bzip68 membrane type
			Glyma.08g360200	NAC domain containing protein 83	Chlorophyll(ide) b reductase
			Glyma.08g360400	Sugar efflux transporter for intercellular exchange	12-oxophytodienoic acid reductase
			Glyma.08g360500	Sugar efflux transporter for intercellular exchange	12-oxophytodienoic acid reductase
			Glyma.08g360600	Serine-Threonine protein kinase	Leucine-rich repeat disease resistance protein
			Glyma.08g360700	Auxin responsive protein	-
			Glyma.08g360800	Pollen Ole e 1 allergen and extensin family protein	Aspartyl protease aed3
			Glyma.08g360900	Signal peptide peptidase	Presenilin
			Glyma.08g361000	Ras family	Vacuolar protein sorting-associated protein 9a
			Glyma.08g361100	RmlC-like cupins superfamily protein	20S proteasome subunit alpha 4
			Glyma.08g361200	Sugar (and other) transporter	Sucrose transport protein SUC3 isoform
ss715613179	12	5190833..5486355	*Glyma.12g071000*	Auxin response factor 4	Eukaryotic translation initiation factor 2c
			Glyma.12g071100	2-methylacyl-CoA dehydrogenase	Methylcrotonoyl-coa carboxylase beta chain
			Glyma.12g071200	Cytochrome B561	Anaphase-promoting complex subunit 4
			Glyma.12g071300	SUMO transferase activity	Spl-ring and pyr_redox_2 domain-containing protein
			Glyma.12g071400	Tetratricopeptide repeat (TPR)-like superfamily protein	Pentatricopeptide repeat-containing protein at1g02370
			Glyma.12g071600	RmlC-like cupins superfamily protein	-
			Glyma.12g071700	Mitochondrial/chloroplast ribosomal protein L28	50S ribosomal protein L27
			Glyma.12g071800	Soluble epoxide hydrolase	Glutamate--trna ligase
			Glyma.12g071900	Zinc finger (C3HC4-type RING finger) family protein	Probable e3 ubiquitin-protein ligase ari8
			Glyma.12g072000	LOB domain-containing protein 1	Ethylene-responsive transcription factor erf021
			Glyma.12g072100	Tetraspanin family	Probable receptor-like protein kinase at5g24010
			Glyma.12g072200	ROP interactive partner 5	Kinesin-like protein kin-13a
			Glyma.12g072300	Ribosomal protein S12/S23 family protein	40S ribosomal protein S29 isoform X2
			Glyma.12g072400	Dof-type zinc finger DNA-binding family protein	Transcription factor bhlh95
			Glyma.12g072500	Arogenate dehydratase	Histidinol-phosphate aminotransferase
			Glyma.12g072600	Sentrin/SUMO-specific protease	Ubiquitin carboxyl-terminal hydrolase l3
			Glyma.12g072700	Vaculolar proton ATPases	V-type h+-transporting atpase subunit a

			Glyma.12g072800	Glycolysis II (from fructose 6-phosphate)	-
			Glyma.12g073000	Mitogen-activated protein kinase	Protein phosphatase 2c and cyclic nucleotide-binding/kinase domain-containing protein isoform x1
			Glyma.12g073100	Ascorbate glutathione cycle	Glutathione dehydrogenase/transferase
			Glyma.12g073200	MOT2 transcription factor	General negative regulator of transcription subunit 3 isoform x1
			Glyma.12g073300	Regulation of transcription	Jasmonate zim domain-containing protein
ss715627820	17	41098767..41215021	*Glyma.17g256900*	Plant protein of unknown function	Trna wybutosine-synthesizing protein 2/3/4 isoform x1
			Glyma.17g257100	Exoribonucleases producing 5'-phosphomonoesters	Decapping nuclease dxo homolog
			Glyma.17g257200	Putative lysine decarboxylase family protein	Bifunctional dihydrofolate reductase-thymidylate synthase
			Glyma.17g257400	DnaJ/Hsp40 cysteine-rich domain superfamily protein	Protein maintenance of psii under high light 1
			Glyma.17g257500	HVA22/DP1 gene product-related proteins	Ubiquitin-conjugating enzyme e2 g1
			Glyma.17g257600	tRNA (adenine-N(1)-)-methyltransferase non-catalytic subunit	Trna (adenine57-n1/adenine58-n1)-methyltransferase catalytic subunit
			Glyma.17g257800	Hexokinase	Glucose-6-phosphate isomerase
			Glyma.17g257700	C_2H_2-like zinc finger protein	Kinetochore-associated protein knl-2 homolog isoform x1
			Glyma.17g257900	Clathrin adaptor complex small chain	AP-2 complex subunit mu-1
			Glyma.17g258100	Gibberellin regulated protein	Xyloglucan endotransglucosylase/hydrolase
ss715627853	17	41433652..41457489	*Glyma.17g260100*	Alcohol dehydrogenase GroES-like domain	Phosphoglycerate kinase
			Glyma.17g260200	Protein of unknown function	Pathogen-related protein
			Glyma.17g260300	Peroxidase	Glycosyltransferases
			Glyma.17g260400	Rare lipoprotein A (RlpA)-like double-psi beta-barrel	Polygalacturonase precursor
ss715636857	20	1655457..1965560	*Glyma.20g017200*	Mitochondrial chaperonin	Chaperonin cpn60-2, mitochondrial isoform x1
			Glyma.20g017400	GDP-fucose protein O-fucosyltransferase	O-fucosyltransferase 23 isoform
			Glyma.20g017500	Phytanoyl-CoA dioxygenase	Fumarate hydratase 1 / Acetyl-coa acyltransferase 1; 3-ketoacyl-CoA thiolase
			Glyma.20g017600	Homeobox domain	Transcription factor yabby13;
			Glyma.20g017700	40S ribosomal protein S17	40S ribosomal protein S29 isoform X2
			Glyma.20g017800	Integral component of membrane	CLAVATA3/ESR (CLE)-related protein 13
			Glyma.20g017900	Pathogenesis-related protein Bet v I family	Polygalacturonase inhibiting protein
			Glyma.20g018000	GDP-glucose biosynthesis	UTP--glucose-1-phosphate uridylyltransferase isoform X2 / Alpha-1,4 glucan phosphorylase L isozyme
			Glyma.20g018100	C2 domain	Lactoylglutathione lyase
			Glyma.20g018200	Glycosyl hydrolases family 17	Glycosyltransferases
			Glyma.20g018300	Succinate:Quinone oxidoreductase	Putative succinate dehydrogenase complex protein
			Glyma.20g018400	Human Cytomegalovirus UL139 protein	-

				Glyma.20g018500	Calcyclin-binding protein CacyBP	Activator of 90 kda heat shock protein atpase homolog isoform x2
				Glyma.20g018600	Cytochrome P450 CYP2 subfamily	Mfs transporter, sp family, sugar:h+ symporter
				Glyma.20g018700	V-Type proton ATPase proteolipid subunit	V-type h+-transporting atpase subunit d
				Glyma.20g018800	Cytochrome P450 CYP2 subfamily	Mfs transporter, sp family, sugar:h+ symporter
				Glyma.20g018900	Cytochrome P450 CYP2 subfamily	Mfs transporter, sp family, sugar:h+ symporter
	ss715637428	20	34276503..34314607	*Glyma.20g099600*	Methionyl-tRNA synthetase	Leucine--tRNA ligase
				Glyma.20g099700	Domain of unknown function	Casparian strip membrane protein
				Glyma.20g099800	Zinc finger, C3HC4 type (RING finger)	Dehydrin
				Glyma.20g099900	Plant protein of unknown function	5-O-(4-coumaroyl)-D-quinate 3'-monooxygenase
Stachyose	ss715583079	2	44214908..44306758	*Glyma.02g255100*	Glycosyl transferase family 64 domain	Xyloglucan galactosyltransferase
				Glyma.02g255200	Tetratricopeptide repeat (TPR)-like superfamily protein	-
				Glyma.02g255300	Integral component of membrane	-
				Glyma.02g255400	Hs1pro-1 N-terminus	O-fucosyltransferase 20
				Glyma.02g255500	Ribosomal protein L10	Large subunit ribosomal protein l3e
				Glyma.02g255600	Ribosomal protein L10	Large subunit ribosomal protein l3e
				Glyma.02g255700	Syntaxin	Golgi snap receptor complex member 2
				Glyma.02g255800	Transmembrane amino acid transporter protein	Solute carrier family 12 (potassium/chloride transporter)
				Glyma.02g255900	Ubiquitin-binding WIYLD domain	-
				Glyma.02g256000	Peptidase family M50	Atp-dependent clp protease proteolytic subunit-related protein 1
				Glyma.02g256100	Leucine-rich repeat protein kinase family protein	U11/U12 small nuclear ribonucleoprotein 20 kDa protein
				Glyma.02g256200	Leucine-rich repeat protein kinase family protein	-
				Glyma.02g256300	Pleckstrin homology (PH) domain superfamily protein	COP1-interactive protein 1
				Glyma.02g256500	Aspartyl protease	Mitochondrial chaperone bcs1
	ss715583119	2	44409157..44553129	*Glyma.02g257600*	Protein kinase domain	Protein phosphatase 2c and cyclic nucleotide-binding/kinase domain-containing protein isoform x1
				Glyma.02g257700	Anthocyanidin synthase	Anthocyanidin reductase 1
				Glyma.02g257800	Synaptobrevin	Vacuolar protein-sorting-associated protein 33 homolog
				Glyma.02g257900	Gibberellin 2β-dioxygenase	Protein brassinosteroid insensitive
				Glyma.02g258000	Response to freezing	Glycosyltransferases
				Glyma.02g258100	Anaphase-promoting complex (APC)	Anaphase-promoting complex subunit
				Glyma.02g258200	Clathrin propeller repeat	Phosphatidylinositol-3,4,5-trisphosphate 3-phosphatase
				Glyma.02g258300	Helix-loop-helix DNA-binding domain	Transcription factor
	ss715592340	5	2167334..2213518	*Glyma.05g024700*	integral component of membrane	-
				Glyma.05g024800	Galactosyl transferase	Xyloglucan galactosyltransferase

			Glyma.05g024900	PPR repeat	50S ribosomal protein L21
			Glyma.05g025000	Response regulator receiver domain	Protein early flowering 4a
			Glyma.05g025100	Exonuclease	Acyl-acyl carrier protein thioesterase atl3
			Glyma.05g025200	Zinc finger, C3HC4 type (RING finger)	Rna-dependent rna polymerase 2 isoform x1
			Glyma.05g025300	Ribulose-phosphate 3 epimerase family	Probable ribose-5-phosphate isomerase
			Glyma.05g025400	Xyloglucan fucosyltransferase	Xyloglucan galactosyltransferase
ss715592442	5	2283415..2372200	*Glyma.05g026200*	Pyridine nucleotide-disulphide oxidoreductase	Chlorophyll/bacteriochlorophyll a synthase
			Glyma.05g026300	Malate dehydrogenase	Malate synthase, glyoxysomal; Malate synthase; Citrate synthase, Fumarate hydratase 1
			Glyma.05g026400	PPR repeat	28S ribosomal protein
			Glyma.05g026500	RWP-RK domain	Tetratricopeptide repeat protein 1
			Glyma.05g026600	Protein of unknown function	60S ribosome subunit biogenesis protein NIP7 homolog
			Glyma.05g026800	Xylogalacturonan β-1,3-xylosyltransferase	Glycosyltransferase family 64 protein c4
			Glyma.05g026900	ZF-HD protein dimerisation region	Homeobox-leucine zipper protein athb-12
			Glyma.05g027000	Myb-like DNA-binding domain	Probable wrky transcription factor 49
			Glyma.05g027100	alpha/beta-Hydrolases superfamily protein	Sucrose synthase
			Glyma.05g027300	Nucleosome assembly protein	Acidic leucine-rich nuclear phosphoprotein 32
ss715591198	5	35722019..35773064	*Glyma.05g166700*	Protein tyrosine kinase	Probable calcium-binding protein cml21 isoform x1
			Glyma.05g166800	Mitochondrial outer membrane protein 25	Fructose-1,6-bisphosphatase
			Glyma.05g166900	Cytochrome P450	Atp-dependent clp protease atp-binding subunit clpb
			Glyma.05g167000	Plant thionin family protein	-
			Glyma.05g167100	Protein of Unknown Function	Dirigent protein 23
ss715601133	8	2300808..2384272	*Glyma.08g028600*	UDP-α-D-galacturonate biosynthesis II (from D-galacturonate)	UDPglucose--hexose-1-phosphate uridylyltransferase / UDP-sugar pyrophosphorylase 1
			Glyma.08g028800	Phospholipase A2 family protein	Lecithin-cholesterol acyltransferase-like 4
			Glyma.08g028900	Type I phosphodiesterase / nucleotide pyrophosphatase	Glutamine-dependent NAD(+) synthetase
			Glyma.08g029000	Cytidine and deoxycytidylate deaminase zinc-binding region	Uridine nucleosidase 1-like
			Glyma.08g029100	Tetraspanin family	Glutathione s-transferase dhar3
			Glyma.08g029200	Ethylene biosynthesis I (plants)	1-aminocyclopropane-1-carboxylate synthase 7-like
			Glyma.08g029300	Sodium/hydrogen exchanger family	Kup system potassium uptake protein
			Glyma.08g029400	Myb-like DNA-binding domain	Transcription factor myb
			Glyma.08g029500	Serine/threonine protein kinase	E3 ubiquitin-protein ligase
			Glyma.08g029600	Putative Phosphatase	Vacuolar protein sorting-associated protein
			Glyma.08g029700	Differentiation-related gene 1 protein	Guanine nucleotide-binding protein

ss715603880	9	3766363..3772351	*Glyma.09g044100*	Uncharacterized conserved protein related to condensin complex subunit 1	Condensin-2 complex subunit H2
ss715639178	9	30134957	*Glyma.09g124300*	Photosynthesis	-
ss715606330	10	3526785..3598580	*Glyma.10g039900*	Ribosomal protein L14p/L23e	Large subunit ribosomal protein l12e
			Glyma.10g040000	Glutathione S-transferase	Alpha-1,3-glucosyltransferase
			Glyma.10g040100	Ubiquitin-binding WIYLD domain	Chromo domain-containing protein lhp1
			Glyma.10g040200	Mitochondrial import receptor subunit TOM22	Mitochondrial import inner membrane translocase subunit
			Glyma.10g040400	AUX/IAA family	Auxin response factor
			Glyma.10g040500	Leucine rich repeat N-terminal domain	Probable lrr receptor-like serine/threonine-protein kinase at5g63710
			Glyma.10g040600	Photosynthesis	Photosystem i reaction center subunit iii
			Glyma.10g040700	Xyloglucan biosynthesis	Xyloglucan 6-xylosyltransferase 2
			Glyma.10g040800	RNA-binding protein	15-cis-phytoene desaturase,
			Glyma.10g040900	Ring zinc finger-containing protein	Ribosomal RNA-processing protein
			Glyma.10g041000	Alcohol dehydrogenase GroES-like domain	S-formylglutathione hydrolase
ss715615716	13	18315025..18531998	*Glyma.13g077500*	Response regulator receiver domain	Btb/poz and taz domain-containing protein 4
			Glyma.13g077600	Protein of unknown function	-
			Glyma.13g077700	Serine hydroxymethyltransferase	Protein phosphatase 2c and cyclic nucleotide-binding/kinase domain-containing protein isoform x1
			Glyma.13g077800	EamA-like transporter family	-
			Glyma.13g077900	Vacuolar H+-ATPase V0 sector, subunits c/c'	V-type h+-transporting atpase subunit d
			Glyma.13g078000	BTB/POZ domain	Cullin-1
			Glyma.13g078200	Leucine Rich Repeat	-
			Glyma.13g078400	Pheophorbide a oxygenase	Chlorophyll(ide) b reductase
			Glyma.13g078500	Plant protein of unknown function	Web family protein at3g51220
			Glyma.13g078600	ATP synthase alpha/beta family	F-type h+-transporting atpase subunit gamma
			Glyma.13g078700	PQ-loop repeat family protein / transmembrane family protein	-
			Glyma.13g078800	Mut7-C RNAse domain	Rna exonuclease 4 isoform x1
ss715614101	13	24090619..24618809	*Glyma.13g128100*	Protein of unknown function (DUF861)	Salt stress-induced hydrophobic peptide esi3
			Glyma.13g128200	Serine/threonine protein kinase	Transcription factor myb
			Glyma.13g128300	Sugar Kinase	DNa repair protein xrcc4
			Glyma.13g128400	Sugar Kinase	DNa repair protein xrcc4
			Glyma.13g128500	Myb-like DNA-binding domain	Dna-directed rna polymerase i subunit rpa49
			Glyma.13g128600	Dolichol phosphate-mannose biosynthesis regulatory protein	Dolichol-phosphate mannosyltransferase subunit 1
			Glyma.13g128700	Myosin ATPase	Kinesin-like protein kin-14i isoform x1
			Glyma.13g128800	Flavin-containing monooxygenase	L-tryptophan--pyruvate aminotransferase 1
			Glyma.13g128900	Glutaredoxin and related proteins	Atp-binding cassette

Glyma.13g129000	Glutathione S-transferase	Glutathione synthetase
Glyma.13g129100	Glutathione S-transferase	Glutathione s-transferase 2 isoform x2
Glyma.13g129200	Plant self-incompatibility protein S1	Phospholipid-transporting ATPase
Glyma.13g129400	Photosystem I reaction centre subunit IV / PsaE	Photosystem i reaction center subunit iii
Glyma.13g129500	Photosystem I reaction centre subunit IV / PsaE	Photosystem i reaction center subunit iii
Glyma.13g129700	Protein of unknown function	-
Glyma.13g129800	Tetraspanin family	Cell division protein ftsz homolog 1
Glyma.13g129900	GDSL-like Lipase/Acylhydrolase	Protein gamete expressed 2
Glyma.13g130000	CLASP N terminal	Cytoskeleton-associated protein 5
Glyma.13g130100	Helix-loop-helix DNA-binding domain	Protein far-red elongated hypocotyl 3-like
Glyma.13g130200	Predicted MYND Zn-finger protein/hormone receptor interactor	Box c/d snorna protein 1
Glyma.13g130300	26S proteasome regulatory complex, subunit PSMD10	Protein phosphatase 2c and cyclic nucleotide-binding/kinase domain-containing protein isoform x1
Glyma.13g130400	ATPase family associated with various cellular activities	Proteasome subunit beta type
Glyma.13g130700	HSP20-like chaperones superfamily protein	Mitochondrial import receptor subunit tom40 homolog 1-like
Glyma.13g130800	HSP20-like chaperones superfamily protein	Mitochondrial import receptor subunit tom40 homolog 1-like
Glyma.13g130900	Heat shock protein 70KDA	Suppressor of tumorigenicity protein 13
Glyma.13g131000	PIF1-like helicase	Dna replication atp-dependent helicase/nuclease dna2 isoform x1
Glyma.13g131100	SANT Associated	L-aminoadipate-semialdehyde dehydrogenase-phosphopantetheinyl transferase isoform x1
Glyma.13g131200	Sodium/hydrogen exchanger family	Na+/h+ antiporter
Glyma.13g131300	Putative GTPase activating protein for Arf	Probable adp-ribosylation factor gtpase-activating protein agd8
Glyma.13g131400	Short chain dehydrogenase	Bifunctional dihydrofolate reductase-thymidylate synthase
Glyma.13g131600	PPR repeat	Guanylate kinase 3
Glyma.13g131700	Cofilin/tropomyosin-type actin-binding protein	Adenylyl cyclase-associated protein
Glyma.13g131800	Acyl-CoA N-acyltransferase with RING/FYVE/PHD-type zinc finger domain	-
Glyma.13g131900	Glutamate N-acetyltransferase	Acetylglutamate kinase
Glyma.13g132100	Histone H2B	Histone acetyltransferase gcn5
Glyma.13g132400	Inorganic diphosphatase	V-type h+-transporting atpase 21kda proteolipid subunit
Glyma.13g132500	EamA-like transporter family	Psbp domain-containing protein 3

			Glyma.13g132700	D-myo-inositol (1,4,5)-trisphosphate degradation	Inositol polyphosphate 6-/3-/5-kinase
			Glyma.13g132800	C_2H_2 and C2HC zinc fingers superfamily protein	Protein phosphatase 2c and cyclic nucleotide-binding/kinase domain-containing protein isoform x1
			Glyma.13g132900	DNA-directed RNA polymerase family protein	-
			Glyma.13g133000	C_2H_2 and C2HC zinc fingers superfamily protein	Protein phosphatase 2c and cyclic nucleotide-binding/kinase domain-containing protein isoform x1
			Glyma.13g133100	C_2H_2-type zinc finger family protein	Protein phosphatase 2c and cyclic nucleotide-binding/kinase domain-containing protein isoform x1
			Glyma.13g133200	Kinesin motor domain	Abnormal spindle-like microcephaly-associated protein homolog isoform x1
			Glyma.13g133300	Plant protein 1589 of unknown function	Ornithine carbamoyltransferase
			Glyma.13g133400	Tetratricopeptide repeat (TPR)-like superfamily protein	U4/U6 small nuclear ribonucleoprotein Prp31 homolog
			Glyma.13g133500	Protein of unknown function	Microtubule-associated protein 70-2
			Glyma.13g133600	Heavy-metal-associated domain	Cytochrome c oxidase assembly protein subunit 17
			Glyma.13g133800	UDP-glucuronate decarboxylase	UDP-glucose 6-dehydrogenase
			Glyma.13g133900	Ribosomal L29 protein	Large subunit ribosomal protein l8e
ss715617675	14	13008001..13358759	*Glyma.14g112400*	Homeobox domain	Transcription repressor ofp5-like
			Glyma.14g112600	Integral component of membrane	Arabinosyltransferase xeg113 isoform x1
			Glyma.14g112700	SET domain	Set1/Ash2 histone methyltransferase complex subunit ASH2
			Glyma.14g112800	Peptidase M16 inactive domain	Aminoacylase-1 isoform X1
			Glyma.14g112900	Triacylglycerol Lipase 1	Diacylglycerol o-acyltransferase 1c
			Glyma.14g113000	ATPase	Abc transporter d family member 1-like isoform x2
			Glyma.14g113100	Polysaccharide binding	Protein phosphatase 2c and cyclic nucleotide-binding/kinase domain-containing protein isoform x1
			Glyma.14g113200	Serine/threonine protein kinase	Protein phosphatase 2c and cyclic nucleotide-binding/kinase domain-containing protein isoform x1
ss715622806	15	51360752..51670112	*Glyma.15g275200*	Thaumatin family	-
			Glyma.15g275300	Triose-phosphate Transporter family	Glucose-1-phosphate adenylyltransferase
			Glyma.15g275400	Glycosyl hydrolases family 28	Pectinesterase
			Glyma.15g275500	Actin	Dna-directed rna polymerase v subunit 1
			Glyma.15g275600	Photosystem II 5 kD protein	Photosystem i subunit
			Glyma.15g275700	Photosystem II 5 kD protein	Photosystem i subunit
			Glyma.15g275900	Protein of unknown function	30S ribosomal protein S1 isoform 1
			Glyma.15g276000	DNA topoisomerase type II	Structural maintenance of chromosomes protein 6b
			Glyma.15g276100	Protein of unknown function	Vacuolar protein sorting-associated protein 29
			Glyma.15g276200	Integral component of membrane	Casein kinase 1-like protein 3
			Glyma.15g276300	Sterol 24-C-methyltransferase	Plant 4,4-dimethylsterol c-4alpha-methyl-monooxygenase
			Glyma.15g276400	Copper amine oxidase	Aldehyde dehydrogenase family 3 member

Glyma.15g276500	Flavin containing amine oxidoreductase	Helicase protein mom1 isoform x1
Glyma.15g276600	Wound-induced protein	-
Glyma.15g276700	Uncharacterized protein family	Glucan endo-1,3-beta-glucosidase a6
Glyma.15g276800	Thiamine pyrophosphate enzyme	2-succinylbenzoate--CoA ligase
Glyma.15g276900	Protein arginine N-methyltransferase	Mediator of rna polymerase ii transcription subunit 36

[1]Gene model: Glycine max cv. Williams 82 reference-genome models version 2.0. [2]Phytozome database. [3]STRING database. [4]Data not available.

Supplementary Table 6. The gene ontology annotations of candidate genes.

Trait	Candidate gene	Category	Gene ontology annotation
Sucrose	*Glyma.02g129200*	GOTERM_MF_DIRECT	GO:0002953 ~ 5'-deoxynucleotidase activity
			GO:0005524 ~ ATP binding
	Glyma.08g356800	GOTERM_BP_DIRECT	GO:0045490 ~ Pectin catabolic process
		GOTERM_MF_DIRECT	GO:0030570 ~ Pectate lyase activity
		GOTERM_MF_DIRECT	GO:0046872 ~ Metal ion binding
		GOTERM_CC_DIRECT	GO:0016020 ~ Membrane
	Glyma.08g357200	GOTERM_MF_DIRECT	GO:0004672 ~ Protein kinase activity
			GO:0005524 ~ ATP binding
	Glyma.08g358700	-[1]	-
	Glyma.08g358800	GOTERM_BP_DIRECT	GO:0006012 ~ Galactose metabolic process
		GOTERM_CC_DIRECT	GO:0005829 ~ Cytosol
		GOTERM_MF_DIRECT	GO:0004335 ~ Galactokinase activity
			GO:0005524 ~ ATP binding
	Glyma.08g360400	GOTERM_BP_DIRECT	GO:0008643 ~ Carbohydrate transport
		GOTERM_CC_DIRECT	GO:0005886 ~ Plasma membrane
			GO:0016020 ~ Membrane
		GOTERM_MF_DIRECT	GO:0051119 ~ Sugar transmembrane transporter activity
	Glyma.08g360500	GOTERM_BP_DIRECT	GO:0008643 ~ Carbohydrate transport
		GOTERM_CC_DIRECT	GO:0005886 ~ Plasma membrane
			GO:0016020 ~ Membrane
		GOTERM_MF_DIRECT	GO:0051119 ~ Sugar transmembrane transporter activity
	Glyma.08g361200	GOTERM_CC_DIRECT	GO:0016020 ~ Membrane
		GOTERM_MF_DIRECT	GO:0005366 ~ Myo-inositol:proton symporter activity
	Glyma.12g072800	GOTERM_BP_DIRECT	GO:0006096 ~ Glycolytic process
		GOTERM_CC_DIRECT	GO:0005737 ~ Cytoplasm
			GO:0009570 ~ Chloroplast stroma
			GO:0016020 ~ Membrane
		GOTERM_MF_DIRECT	GO:0000287 ~ Magnesium ion binding

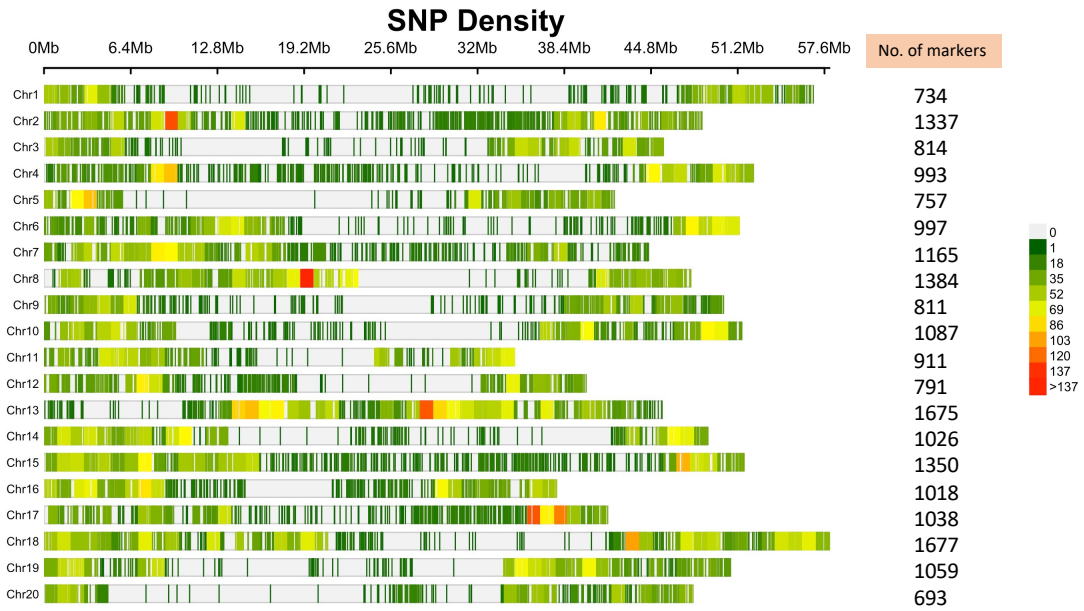
			GO:0004743 ~ Pyruvate kinase activity
			GO:0005524 ~ ATP binding
			GO:0016301 ~ Kinase activity
			GO:0030955 ~ Potassium ion binding
			GO:0001678 ~ Intracellular glucose homeostasis
		GOTERM_BP_DIRECT	GO:0006096 ~ Glycolytic process
			GO:0019318 ~ Hexose metabolic process
			GO:0051156 ~ Glucose 6-phosphate metabolic process
		GOTERM_CC_DIRECT	GO:0005739 ~ Mitochondrion
	Glyma.17g257800		GO:0005829 ~ Cytosol
			GO:0004340 ~ Glucokinase activity
			GO:0005524 ~ ATP binding
		GOTERM_MF_DIRECT	GO:0005536 ~ Glucose binding
			GO:0008865 ~ Fructokinase activity
			GO:0019158 ~ Mannokinase activity
	Glyma.17g258100	-	-
		GOTERM_BP_DIRECT	GO:0006979 ~ Response to oxidative stress
		GOTERM_CC_DIRECT	GO:0009505 ~ Plant-type cell wall
	Glyma.17g260300		GO:0004601 ~ Peroxidase activity
		GOTERM_MF_DIRECT	GO:0020037 ~ Heme binding
			GO:0046872 ~ Metal ion binding
		GOTERM_BP_DIRECT	GO:0009664 ~ Plant-type cell wall organization
	Glyma.17g260400	GOTERM_CC_DIRECT	GO:0005576 ~ Extracellular region
			GO:0016020 ~ Membrane
		GOTERM_BP_DIRECT	GO:0006004 ~ Fucose metabolic process
	Glyma.20g017400	GOTERM_CC_DIRECT	GO:0016020 ~ Membrane
		GOTERM_MF_DIRECT	GO:0016740 ~ Transferase activity
		GOTERM_BP_DIRECT	GO:0005975 ~ Carbohydrate metabolic process
	Glyma.20g018000	GOTERM_MF_DIRECT	GO:0004615 ~ Phosphomannomutase activity
			GO:0046872 ~ Metal ion binding
		GOTERM_BP_DIRECT	GO:0005975 ~ Carbohydrate metabolic process
	Glyma.20g018200	GOTERM_CC_DIRECT	GO:0016020 ~ Membrane
		GOTERM_MF_DIRECT	GO:0042973 ~ Glucan endo-1,3-beta-D-glucosidase activity
		GOTERM_BP_DIRECT	GO:0006431 ~ Methionyl-tRNA aminoacylation
	Glyma.20g099600	GOTERM_CC_DIRECT	GO:0005739 ~ Mitochondrion
		GOTERM_MF_DIRECT	GO:0004825 ~ Methionine-tRNA ligase activity
			GO:0005524 ~ ATP binding
Stachyose		GOTERM_BP_DIRECT	GO:0006486 ~ Protein glycosylation
	Glyma.02g255100	GOTERM_CC_DIRECT	GO:0016020 ~ Membrane
		GOTERM_MF_DIRECT	GO:0016757 ~ Glycosyltransferase activity
	Glyma.02g255400	GOTERM_BP_DIRECT	GO:0006952 ~ Defense response
			GO:0019441 ~ Tryptophan catabolic process to kynurenine

Glyma.02g258000	GOTERM_MF_DIRECT	GO:0020037 ~ Heme binding
		GO:0046872 ~ Metal ion binding
	GOTERM_CC_DIRECT	GO:0005794 ~ Golgi apparatus
		GO:0016020 ~ Membrane
	GOTERM_MF_DIRECT	GO:0016413 ~ O-acetyltransferase activity
Glyma.02g258200	GOTERM_BP_DIRECT	GO:0006886 ~ Intracellular protein transport
		GO:0006898 ~ Receptor-mediated endocytosis
	GOTERM_CC_DIRECT	GO:0030130 ~ Clathrin coat of trans-Golgi network vesicle
		GO:0030132 ~ Clathrin coat of coated pit
		GO:0071439 ~ Clathrin complex
	GOTERM_MF_DIRECT	GO:0005198 ~ Structural molecule activity
		GO:0032051 ~ Clathrin light chain binding
Glyma.05g024800	GOTERM_BP_DIRECT	GO:0009969 ~ Xlyoglucan biosynthetic process
	GOTERM_CC_DIRECT	GO:0000139 ~ Golgi membrane
		GO:0016020 ~ Membrane
	GOTERM_MF_DIRECT	GO:0016758 ~ Hexosyltransferase activity
Glyma.05g025300	GOTERM_BP_DIRECT	GO:0009052 ~ Pentose-phosphate shunt, non-oxidative branch
		GO:0019323 ~ Pentose catabolic process
		GO:0005975 ~ Carbohydrate metabolic process
	GOTERM_CC_DIRECT	GO:0005829 ~ Cytosol
	GOTERM_MF_DIRECT	GO:0004750 ~ D-ribulose-phosphate 3-epimerase activity
		GO:0046872 ~ Metal ion binding
Glyma.05g025400	GOTERM_BP_DIRECT	GO:0009969 ~ Xlyoglucan biosynthetic process
		GO:0071555 ~ Cell wall organization
	GOTERM_CC_DIRECT	GO:0016020 ~ Membrane
		GO:0032580 ~ Golgi cisterna membrane
	GOTERM_MF_DIRECT	GO:0008107 ~ Galactoside 2-alpha-L-fucosyltransferase activity
		GO:0008417 ~ Fucosyltransferase activity
Glyma.05g026800	GOTERM_BP_DIRECT	GO:0006486 ~ Protein glycosylation
	GOTERM_CC_DIRECT	GO:0000139 ~ Golgi membrane
		GO:0016020 ~ Membrane
	GOTERM_MF_DIRECT	GO:0016757 ~ Glycosyltransferase activity
Glyma.05g027100	-	-
Glyma.05g166800	-	-
Glyma.08g028600	GOTERM_BP_DIRECT	GO:0006012 ~ Galactose metabolic process
	GOTERM_CC_DIRECT	GO:0005737 ~ Cytoplasm
		GO:0005829 ~ Cytosol

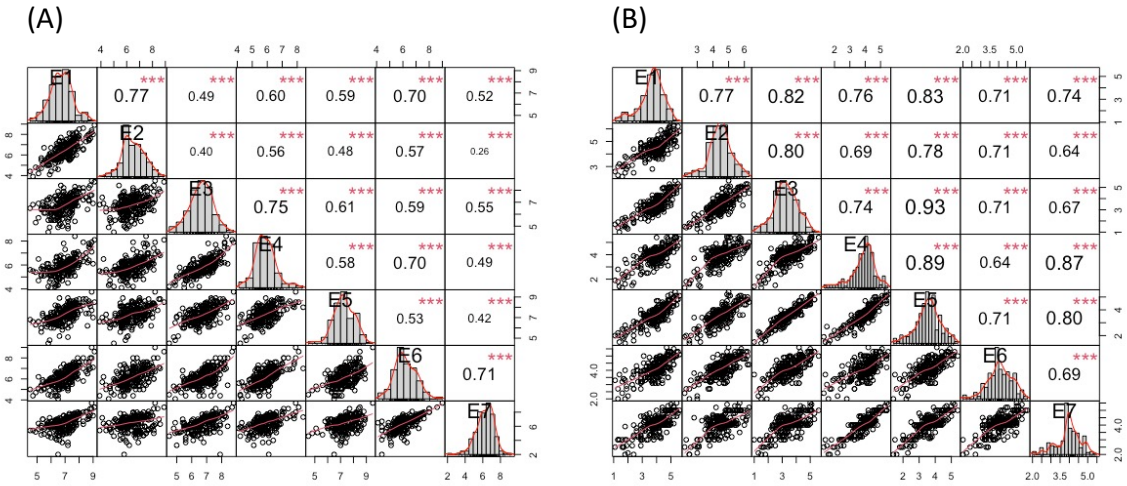
	GOTERM_MF_DIRECT	GO:0004335 ~ Galactokinase activity
		GO:0005524 ~ ATP binding
		GO:0047912 ~ Galacturonokinase activity
	GOTERM_BP_DIRECT	GO:0007076 ~ Mitotic chromosome condensation
		GO:0010032 ~ Meiotic chromosome condensation
Glyma.09g044100	GOTERM_CC_DIRECT	GO:0000779 ~ Condensed chromosome, centromeric region
		GO:0005634 ~ Nucleus
	GOTERM_MF_DIRECT	GO:0042393 ~ Histone binding
	GOTERM_BP_DIRECT	GO:0015979 ~ Photosynthesis
Glyma.09g124300	GOTERM_CC_DIRECT	GO:0009535 ~ Chloroplast thylakoid membrane
		GO:0009538 ~ Photosystem I reaction center
		GO:0016020 ~ Membrane
	GOTERM_BP_DIRECT	GO:0006749 ~ Glutathione metabolic process
Glyma.10g040000	GOTERM_CC_DIRECT	GO:0005737 ~ Cytoplasm
	GOTERM_MF_DIRECT	GO:0004364 ~ Glutathione transferase activity
	GOTERM_BP_DIRECT	GO:0006486 ~ Protein glycosylation
Glyma.10g040700	GOTERM_CC_DIRECT	GO:0000139 ~ Golgi membrane
		GO:0016020 ~ Membrane
	GOTERM_MF_DIRECT	GO:0016757 ~ Glycosyltransferase activity
Glyma.13g077900	GOTERM_CC_DIRECT	GO:0005774 ~ Vacuolar membrane
		GO:0016020 ~ Membrane
		GO:0033179 ~ Proton-transporting V-type ATPase, V0 domain
	GOTERM_MF_DIRECT	GO:0015078 ~ Proton transmembrane transporter activity
Glyma.13g128300	GOTERM_BP_DIRECT	GO:0019252 ~ Starch biosynthetic process
		GO:0044281 ~ Small molecule metabolic process
	GOTERM_MF_DIRECT	GO:0016773 ~ phosphotransferase activity, alcohol group as acceptor
		GO:0019200 ~ Carbohydrate kinase activity
Glyma.13g128400	GOTERM_BP_DIRECT	GO:0071704 ~ Organic substance metabolic process
	GOTERM_MF_DIRECT	GO:0016301 ~ Kinase activity
		GO:0016773 ~ phosphotransferase activity, alcohol group as acceptor
Glyma.13g132700	GOTERM_BP_DIRECT	GO:0046856 ~ Phosphatidylinositol dephosphorylation
	GOTERM_MF_DIRECT	GO:0004439 ~ Phosphatidylinositol-4,5-bisphosphate 5-phosphatase activity
		GO:0034485 ~ Phosphatidylinositol-3,4,5-trisphosphate 5-phosphatase activity
Glyma.13g133800	GOTERM_BP_DIRECT	GO:0033320 ~ UDP-D-xylose biosynthetic process
		GO:0042732 ~ D-xylose metabolic process
	GOTERM_CC_DIRECT	GO:0005737 ~ Cytoplasm
		GO:0016020 ~ Membrane

Glyma.14g113100	GOTERM_MF_DIRECT	GO:0048040 ~ UDP-glucuronate decarboxylase activity
		GO:0070403 ~ NAD+ binding
	GOTERM_CC_DIRECT	GO:0016020 ~ Membrane
	GOTERM_MF_DIRECT	GO:0030247 ~ Polysaccharide binding
Glyma.15g275300	GOTERM_CC_DIRECT	GO:0005794 ~ Golgi apparatus
		GO:0016020 ~ Membrane
		GO:0031969 ~ Chloroplast membrane
	GOTERM_MF_DIRECT	GO:0015120 ~ Phosphoglycerate transmembrane transporter activity
		GO:0015297 ~ Antiporter activity
		GO:0015605 ~ Organophosphate ester transmembrane transporter activity
Glyma.15g275400	GOTERM_BP_DIRECT	GO:0005975 ~ Carbohydrate metabolic process
	GOTERM_MF_DIRECT	GO:0004650 ~ Polygalacturonase activity
Glyma.15g276700	GOTERM_CC_DIRECT	GO:0009706 ~ Chloroplast inner membrane
		GO:0016020 ~ Membrane

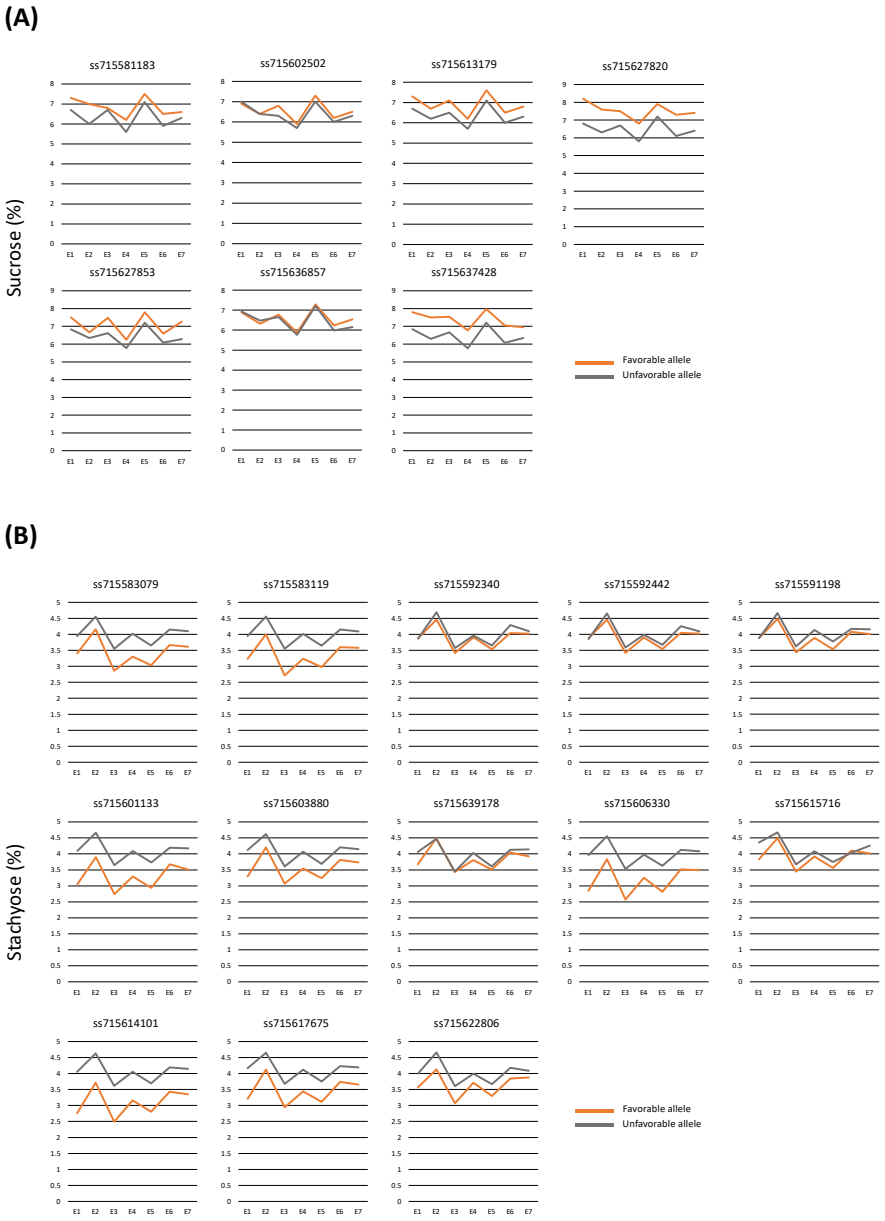
[1]Data not available



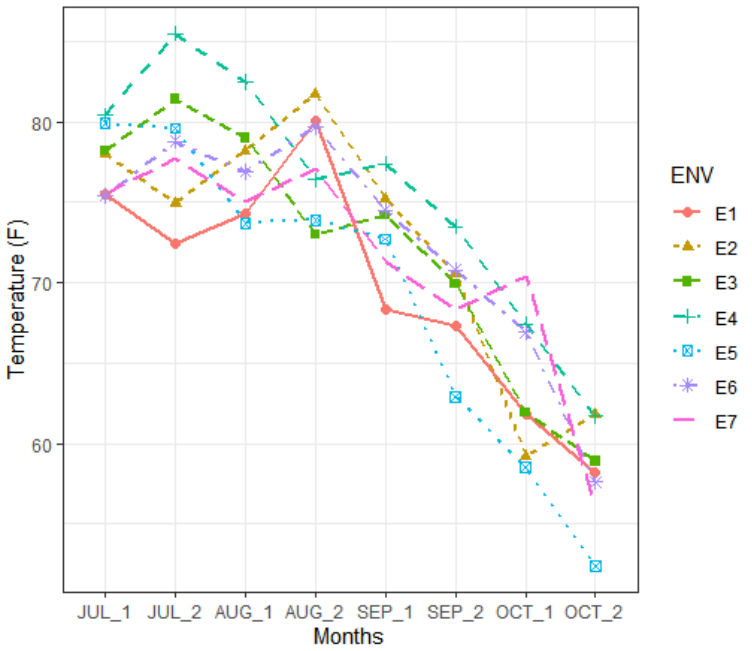
Supplementary Figure 1. The number of filtered SNPs mapped across 20 soybean chromosomes.



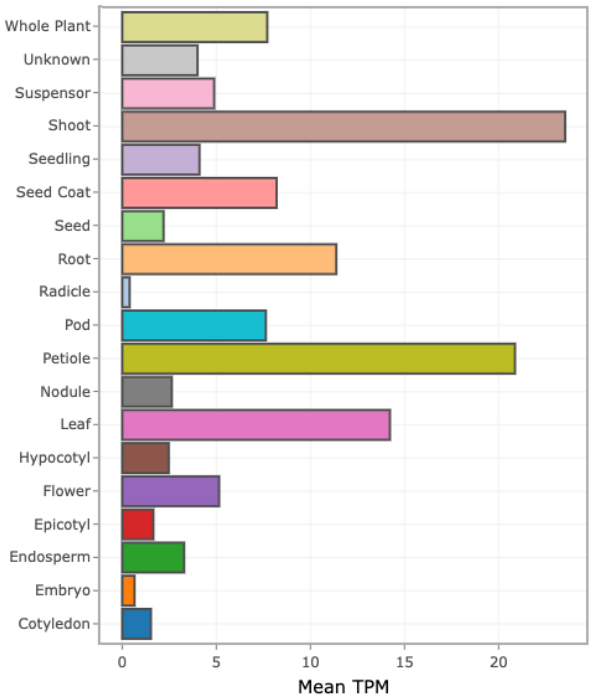
Supplementary Figure 2. Pearson's correlations among environments for sucrose (A) and stachyose (B) content. E1, Fayetteville, AR, 2014; E2, Stuttgart, AR, 2014; E3, Fayetteville, AR, 2015; E4, Stuttgart, AR, 2015; E5, Portageville, MO, 2020; E6, Columbia, MO, 2021; E7, Portageville, MO, 2021.



Supplementary Figure 3. Allelic effect of significant SNPs for sucrose (A) and stachyose (B) content across environments. The allele conferring higher sucrose content was designated a favorable allele, while the allele conferring lower stachyose content was designated a favorable allele. E1, Fayetteville, AR, 2014; E2, Stuttgart, AR, 2014; E3, Fayetteville, AR, 2015; E4, Stuttgart, AR, 2015; E5, Portageville, MO, 2020; E6, Columbia, MO, 2021; E7, Portageville, MO, 2021.



Supplementary Figure 4. Temperatures during soybean pod-filling stages (July – October) across environments. '_1' and '_2' indicate the first 15 days and the rest of the month, respectively. E1, Fayetteville, AR, 2014; E2, Stuttgart, AR, 2014; E3, Fayetteville, AR, 2015; E4, Stuttgart, AR, 2015; E5, Portageville, MO, 2020; E6, Columbia, MO, 2021; E7, Portageville, MO, 2021.



Supplementary Figure 5. The gene expression level of raffinose synthase 3 gene (*Glyma.05g003900*, Wm82.a2.v1) across 19 different parts of the soybean plant. The raw gene expression counts were normalized using a Transcripts Per Million (TPM) methods.

Supplementary tables and figures for Chapter 3

Supplementary Table 1. Analysis of variance for two mapping populations, P593_RIL1 and P593_RIL2.

Source	P593_RIL1				P593_RIL2			
	df	Mean Square	F-value	Pr > F	df	Mean Square	F-value	Pr > F
Genotype	139	2.5	4.5	< .0001	206	2.9	5.6	< .0001
Environment	2	79.6	16.4	0.0253	2	154.1	22.6	0.0142
Genotype × Environment	258	0.6	1.0	0.4515	406	0.5	1.4	0.0005
Replication (Environment)	3	4.5	8.3	< .0001	3	6.4	16.6	< .0001
Residual	259	0.5	-	-	476	0.4	-	-

Supplementary Table 2. A list of unique nonsynonymous SNP variants identified in 11 sucrose-related QTL in PI 506593 on Wm82.a2.v1 reference genome assembly.

Chr	Gene ID Wm82.a2.v1	Position	Reference	Variant	Exon
Chr04	*Glyma.04G196400*	46833272	T	G	exon2
Chr04	*Glyma.04G203600*	47633734	G	A	exon4
Chr04	*Glyma.04G204400*	47667694	T	G	exon6
Chr04	*Glyma.04G205000*	47757425	G	C	exon3
Chr04	*Glyma.04G205000*	47757534	G	A	exon3
Chr04	*Glyma.04G205000*	47757561	G	A	exon3
Chr04	*Glyma.04G205000*	47758055	A	G	exon3
Chr04	*Glyma.04G205000*	47758082	C	T	exon3
Chr04	*Glyma.04G205500*	47813697	T	C	exon7
Chr04	*Glyma.04G206000*	47860066	G	T	exon4
Chr04	*Glyma.04G206100*	47873782	A	T	exon6
Chr04	*Glyma.04G206400*	47894201	T	A	exon10
Chr04	*Glyma.04G206400*	47897379	C	T	exon8
Chr04	*Glyma.04G206400*	47900156	A	T	exon4
Chr04	*Glyma.04G206400*	47900654	A	G	exon2
Chr04	*Glyma.04G206400*	47901785	T	C	exon1
Chr04	*Glyma.04G206700*	47924463	T	G	exon1
Chr04	*Glyma.04G206700*	47925101	G	A	exon1
Chr05	*Glyma.05G149500*	34370281	A	T	exon5
Chr05	*Glyma.05G149500*	34370796	T	C	exon4
Chr05	*Glyma.05G149500*	34371162	G	C	exon3
Chr05	*Glyma.05G149500*	34372629	G	A	exon1
Chr05	*Glyma.05G149600*	34374545	G	A	exon4
Chr05	*Glyma.05G149600*	34374840	A	G	exon4
Chr05	*Glyma.05G149700*	34378331	T	C	exon12
Chr05	*Glyma.05G149700*	34382769	T	C	exon3
Chr05	*Glyma.05G149700*	34383187	G	A	exon1
Chr05	*Glyma.05G150800*	34458339	T	A	exon1
Chr05	*Glyma.05G153700*	34713331	G	T	exon2
Chr05	*Glyma.05G156000*	34868956	A	G	exon4
Chr05	*Glyma.05G156000*	34870969	A	T	exon3
Chr05	*Glyma.05G156400*	34889638	A	C	exon4
Chr05	*Glyma.05G156500*	34894175	A	C	exon2
Chr05	*Glyma.05G156500*	34894216	T	A	exon2
Chr05	*Glyma.05G156500*	34894468	C	T	exon2
Chr05	*Glyma.05G156500*	34894523	T	A	exon2
Chr05	*Glyma.05G156500*	34894817	C	G	exon2
Chr05	*Glyma.05G156700*	34913639	A	G	exon1
Chr06	*Glyma.06G191300*	16808865	T	A	exon10
Chr06	*Glyma.06G191400*	16824224	T	C	exon4

Chr06	*Glyma.06G199500*	18213510	T	A	exon4
Chr06	*Glyma.06G200200*	18329608	A	C	exon4
Chr06	*Glyma.06G200400*	18348596	C	T	exon1
Chr06	*Glyma.06G201000*	18415745	T	C	exon2
Chr06	*Glyma.06G201000*	18416737	A	C	exon1
Chr06	*Glyma.06G201700*	18579478	C	G	exon4
Chr06	*Glyma.06G201700*	18579487	G	C	exon4
Chr06	*Glyma.06G201700*	18579715	A	T	exon5
Chr06	*Glyma.06G201700*	18580318	C	T	exon6
Chr06	*Glyma.06G201700*	18587822	G	A	exon15
Chr06	*Glyma.06G201700*	18607360	A	G	exon22
Chr06	*Glyma.06G202400*	18786083	A	G	exon2
Chr06	*Glyma.06G204400*	19251815	C	G	exon3
Chr06	*Glyma.06G204600*	19310302	C	T	exon6
Chr06	*Glyma.06G204600*	19315012	G	A	exon3
Chr06	*Glyma.06G204700*	19317476	G	T	exon1
Chr06	*Glyma.06G204700*	19317477	C	T	exon1
Chr06	*Glyma.06G204700*	19317602	A	G	exon1
Chr06	*Glyma.06G204700*	19318118	T	G	exon1
Chr06	*Glyma.06G204700*	19318226	C	T	exon1
Chr06	*Glyma.06G204700*	19318236	T	C	exon1
Chr06	*Glyma.06G204700*	19318524	A	G	exon1
Chr06	*Glyma.06G206100*	19720911	T	C	exon5
Chr06	*Glyma.06G206300*	19781043	C	T	exon3
Chr06	*Glyma.06G207300*	20012169	G	A	exon3
Chr06	*Glyma.06G208700*	20399638	A	G	exon2
Chr06	*Glyma.06G215300*	22292328	T	C	exon3
Chr06	*Glyma.06G215500*	22375370	A	T	exon3
Chr06	*Glyma.06G216000*	22475961	A	C	exon2
Chr06	*Glyma.06G216100*	22489636	G	A	exon2
Chr06	*Glyma.06G222200*	28895070	G	T	exon1
Chr06	*Glyma.06G226400*	32865236	C	T	exon1
Chr06	*Glyma.06G230700*	36002459	G	A	exon2
Chr06	*Glyma.06G232900*	36962740	G	C	exon2
Chr06	*Glyma.06G233000*	36970778	T	G	exon6
Chr06	*Glyma.06G236500*	38498421	C	A	exon1
Chr06	*Glyma.06G237800*	38799086	C	A	exon3
Chr06	*Glyma.06G242000*	40079197	G	T	exon2
Chr06	*Glyma.06G242000*	40079198	A	C	exon2
Chr06	*Glyma.06G243800*	40570995	G	A	exon17
Chr06	*Glyma.06G243800*	40571037	A	T	exon17
Chr06	*Glyma.06G249300*	42121573	G	T	exon2
Chr06	*Glyma.06G250500*	42391664	T	A	exon4
Chr06	*Glyma.06G251800*	42758837	G	T	exon1
Chr06	*Glyma.06G256000*	43771762	G	A	exon2
Chr06	*Glyma.06G256000*	43771891	A	T	exon2
Chr06	*Glyma.06G256800*	43910001	C	T	exon2
Chr06	*Glyma.06G257400*	44043987	C	A	exon1
Chr06	*Glyma.06G258800*	44378312	G	A	exon1
Chr06	*Glyma.06G258900*	44382527	G	T	exon1
Chr06	*Glyma.06G258900*	44382567	C	T	exon1
Chr06	*Glyma.06G258900*	44382785	C	A	exon1
Chr06	*Glyma.06G258900*	44382886	T	A	exon1
Chr06	*Glyma.06G258900*	44382911	G	T	exon1
Chr06	*Glyma.06G258900*	44382963	A	G	exon1
Chr06	*Glyma.06G258900*	44383992	A	T	exon2
Chr06	*Glyma.06G258900*	44384072	G	A	exon2
Chr06	*Glyma.06G258900*	44384309	C	G	exon3
Chr06	*Glyma.06G258900*	44384372	A	C	exon3
Chr06	*Glyma.06G258900*	44384619	T	C	exon4
Chr06	*Glyma.06G258900*	44384930	T	A	exon5
Chr06	*Glyma.06G258900*	44385014	G	A	exon5

Chr06	*Glyma.06G258900*	44385030	G	A	exon5
Chr06	*Glyma.06G258900*	44385496	G	A	exon7
Chr06	*Glyma.06G258900*	44385621	T	A	exon7
Chr06	*Glyma.06G258900*	44385673	A	C	exon7
Chr06	*Glyma.06G259100*	44411010	G	C	exon6
Chr06	*Glyma.06G259100*	44411022	T	C	exon6
Chr06	*Glyma.06G259100*	44411041	G	T	exon6
Chr06	*Glyma.06G259100*	44411061	G	C	exon6
Chr06	*Glyma.06G259100*	44411092	C	G	exon6
Chr06	*Glyma.06G259100*	44411226	C	T	exon5
Chr06	*Glyma.06G259100*	44411257	T	C	exon5
Chr06	*Glyma.06G259100*	44412707	C	T	exon3
Chr06	*Glyma.06G259200*	44418570	T	G	exon3
Chr06	*Glyma.06G259300*	44427995	C	T	exon2
Chr06	*Glyma.06G259400*	44428529	A	T	exon5
Chr06	*Glyma.06G259400*	44435675	A	C	exon6
Chr06	*Glyma.06G259400*	44435733	C	T	exon6
Chr06	*Glyma.06G259400*	44435752	T	A	exon6
Chr06	*Glyma.06G259400*	44435852	G	C	exon6
Chr06	*Glyma.06G259400*	44435882	A	G	exon6
Chr06	*Glyma.06G259400*	44435948	A	C	exon6
Chr06	*Glyma.06G259400*	44435988	G	T	exon6
Chr06	*Glyma.06G259400*	44437033	C	T	exon4
Chr06	*Glyma.06G259400*	44437944	C	T	exon2
Chr06	*Glyma.06G259400*	44438682	A	G	exon2
Chr06	*Glyma.06G259800*	44468234	T	C	exon5
Chr06	*Glyma.06G259900*	44474218	G	A	exon2
Chr06	*Glyma.06G260100*	44521526	C	A	exon4
Chr06	*Glyma.06G260100*	44521636	T	A	exon4
Chr06	*Glyma.06G260100*	44521638	T	C	exon4
Chr06	*Glyma.06G260100*	44521774	C	G	exon4
Chr06	*Glyma.06G260100*	44521778	T	G	exon4
Chr06	*Glyma.06G260100*	44521844	T	A	exon4
Chr06	*Glyma.06G260100*	44521848	T	G	exon4
Chr06	*Glyma.06G260100*	44521993	A	T	exon4
Chr06	*Glyma.06G260100*	44523588	C	T	exon2
Chr06	*Glyma.06G260100*	44524368	C	T	exon1
Chr06	*Glyma.06G260100*	44524369	T	C	exon1
Chr06	*Glyma.06G260100*	44524465	A	C	exon1
Chr06	*Glyma.06G260100*	44524477	C	G	exon1
Chr06	*Glyma.06G260100*	44524554	C	T	exon1
Chr06	*Glyma.06G260100*	44524807	C	G	exon1
Chr06	*Glyma.06G260600*	44582308	G	A	exon2
Chr06	*Glyma.06G260900*	44665324	A	G	exon1
Chr06	*Glyma.06G260900*	44665595	C	T	exon1
Chr06	*Glyma.06G260900*	44665650	C	A	exon1
Chr06	*Glyma.06G261000*	44684291	C	T	exon1
Chr06	*Glyma.06G261100*	44724382	G	A	exon1
Chr06	*Glyma.06G261100*	44724439	T	A	exon1
Chr06	*Glyma.06G261100*	44724550	C	T	exon1
Chr06	*Glyma.06G261400*	44788868	A	G	exon8
Chr06	*Glyma.06G261400*	44789542	G	C	exon7
Chr06	*Glyma.06G261500*	44830359	G	C	exon4
Chr06	*Glyma.06G261500*	44830369	C	T	exon4
Chr06	*Glyma.06G261500*	44830432	A	C	exon4
Chr06	*Glyma.06G261500*	44832541	G	A	exon2
Chr06	*Glyma.06G261500*	44832961	C	G	exon1
Chr06	*Glyma.06G261500*	44833078	A	T	exon1
Chr06	*Glyma.06G261500*	44833083	G	C	exon1
Chr06	*Glyma.06G261500*	44833325	C	T	exon1
Chr06	*Glyma.06G263800*	45126172	G	A	exon3
Chr06	*Glyma.06G263800*	45126178	T	C	exon3

Chr06	*Glyma.06G263800*	45126366	C	T	exon4
Chr06	*Glyma.06G267000*	45499302	T	C	exon6
Chr06	*Glyma.06G267000*	45500186	C	T	exon5
Chr06	*Glyma.06G267000*	45501678	C	T	exon2
Chr06	*Glyma.06G267000*	45501801	T	C	exon2
Chr06	*Glyma.06G267100*	45514201	T	C	exon2
Chr06	*Glyma.06G267300*	45555550	G	A	exon1
Chr06	*Glyma.06G267300*	45555602	T	G	exon1
Chr06	*Glyma.06G267300*	45555668	A	C	exon1
Chr06	*Glyma.06G267300*	45555829	T	A	exon2
Chr06	*Glyma.06G267300*	45555840	A	G	exon2
Chr06	*Glyma.06G267300*	45555927	G	C	exon2
Chr06	*Glyma.06G267300*	45556309	G	A	exon2
Chr06	*Glyma.06G267300*	45556406	A	T	exon2
Chr06	*Glyma.06G267400*	45562607	C	T	exon5
Chr06	*Glyma.06G267400*	45562629	A	T	exon5
Chr06	*Glyma.06G267400*	45562833	A	C	exon5
Chr06	*Glyma.06G267400*	45562985	T	C	exon5
Chr06	*Glyma.06G267400*	45563941	C	T	exon4
Chr06	*Glyma.06G267500*	45611873	T	C	exon3
Chr06	*Glyma.06G267500*	45611923	C	T	exon3
Chr06	*Glyma.06G268000*	45650736	C	A	exon8
Chr06	*Glyma.06G268000*	45650781	C	A	exon8
Chr06	*Glyma.06G268000*	45650837	G	C	exon8
Chr06	*Glyma.06G268600*	45721742	T	A	exon1
Chr06	*Glyma.06G268600*	45721969	C	G	exon1
Chr06	*Glyma.06G268600*	45722169	G	A	exon1
Chr06	*Glyma.06G268600*	45722204	C	T	exon1
Chr06	*Glyma.06G268600*	45722367	G	A	exon2
Chr06	*Glyma.06G268600*	45722918	T	C	exon2
Chr06	*Glyma.06G268600*	45723114	G	C	exon2
Chr06	*Glyma.06G268600*	45723119	A	G	exon2
Chr06	*Glyma.06G268600*	45723186	A	G	exon2
Chr06	*Glyma.06G268600*	45723611	A	G	exon3
Chr06	*Glyma.06G268600*	45723629	T	C	exon3
Chr06	*Glyma.06G268600*	45723659	A	G	exon3
Chr06	*Glyma.06G268600*	45723688	A	C	exon3
Chr06	*Glyma.06G268600*	45723837	C	A	exon3
Chr06	*Glyma.06G268700*	45726815	T	C	exon6
Chr06	*Glyma.06G268700*	45728903	C	G	exon4
Chr06	*Glyma.06G268700*	45728928	A	C	exon4
Chr06	*Glyma.06G271000*	46018262	A	C	exon3
Chr06	*Glyma.06G271000*	46018654	A	C	exon3
Chr06	*Glyma.06G271900*	46203833	C	T	exon1
Chr06	*Glyma.06G272000*	46217528	G	A	exon2
Chr06	*Glyma.06G272200*	46282615	A	T	exon3
Chr06	*Glyma.06G272200*	46282989	T	A	exon3
Chr06	*Glyma.06G272200*	46286778	T	C	exon1
Chr06	*Glyma.06G272300*	46298841	C	T	exon1
Chr06	*Glyma.06G272400*	46320053	G	A	exon2
Chr06	*Glyma.06G273200*	46407708	G	A	exon1
Chr06	*Glyma.06G276000*	46703908	A	G	exon1
Chr06	*Glyma.06G276000*	46704090	C	A	exon1
Chr06	*Glyma.06G276100*	46707483	T	G	exon5
Chr06	*Glyma.06G276100*	46708228	C	T	exon5
Chr06	*Glyma.06G276100*	46708512	T	C	exon5
Chr06	*Glyma.06G276800*	46798685	A	T	exon2
Chr06	*Glyma.06G276800*	46798780	G	C	exon2
Chr06	*Glyma.06G276800*	46799072	C	G	exon2
Chr06	*Glyma.06G276800*	46799961	T	A	exon3
Chr06	*Glyma.06G276800*	46800208	G	A	exon3
Chr06	*Glyma.06G276800*	46800242	A	G	exon3

Chr08	*Glyma.08G292200*	40620161	G	T	exon3
Chr08	*Glyma.08G292200*	40620204	T	C	exon3
Chr08	*Glyma.08G292200*	40620357	C	T	exon2
Chr08	*Glyma.08G292200*	40620543	T	C	exon2
Chr08	*Glyma.08G292200*	40620959	G	A	exon1
Chr08	*Glyma.08G292200*	40620974	C	G	exon1
Chr08	*Glyma.08G293000*	40768008	C	T	exon15
Chr08	*Glyma.08G293000*	40770365	A	G	exon13
Chr08	*Glyma.08G293100*	40799770	C	T	exon4
Chr08	*Glyma.08G293100*	40799779	G	A	exon4
Chr08	*Glyma.08G293500*	40859845	T	G	exon1
Chr08	*Glyma.08G293900*	40892165	C	T	exon19
Chr08	*Glyma.08G294100*	40944672	A	T	exon3
Chr08	*Glyma.08G294100*	40945699	C	A	exon5
Chr08	*Glyma.08G294300*	40956342	C	A	exon5
Chr08	*Glyma.08G294500*	40967908	A	T	exon2
Chr08	*Glyma.08G294500*	40969586	C	A	exon2
Chr08	*Glyma.08G294800*	40994104	A	G	exon1
Chr08	*Glyma.08G295100*	41011629	G	A	exon2
Chr08	*Glyma.08G295100*	41011634	T	A	exon2
Chr08	*Glyma.08G295100*	41011950	C	T	exon1
Chr08	*Glyma.08G295100*	41011951	G	T	exon1
Chr08	*Glyma.08G295200*	41017625	A	G	exon1
Chr08	*Glyma.08G295300*	41018325	G	A	exon1
Chr08	*Glyma.08G295400*	41032973	G	C	exon1
Chr08	*Glyma.08G295700*	41069193	C	G	exon5
Chr08	*Glyma.08G295800*	41101336	G	A	exon3
Chr08	*Glyma.08G295800*	41101393	T	C	exon3
Chr08	*Glyma.08G296100*	41176884	A	T	exon4
Chr08	*Glyma.08G296300*	41217767	C	G	exon4
Chr08	*Glyma.08G296300*	41219747	A	G	exon6
Chr08	*Glyma.08G296500*	41242165	G	A	exon9
Chr08	*Glyma.08G297000*	41328845	A	C	exon8
Chr08	*Glyma.08G297400*	41360513	C	G	exon3
Chr08	*Glyma.08G297400*	41361211	C	T	exon1
Chr08	*Glyma.08G297500*	41379672	C	T	exon3
Chr08	*Glyma.08G297500*	41379875	C	G	exon3
Chr08	*Glyma.08G297600*	41401710	C	T	exon4
Chr08	*Glyma.08G297600*	41401746	A	G	exon4
Chr08	*Glyma.08G297600*	41401774	C	T	exon4
Chr08	*Glyma.08G297600*	41401812	T	C	exon4
Chr08	*Glyma.08G297600*	41401962	C	T	exon4
Chr08	*Glyma.08G297600*	41402058	C	T	exon4
Chr08	*Glyma.08G297600*	41403816	T	C	exon1
Chr08	*Glyma.08G297700*	41412617	A	T	exon1
Chr08	*Glyma.08G297700*	41412685	G	C	exon1
Chr08	*Glyma.08G297700*	41415651	C	T	exon3
Chr08	*Glyma.08G299500*	41767566	T	C	exon4
Chr08	*Glyma.08G299500*	41767569	T	C	exon4
Chr08	*Glyma.08G299500*	41769008	G	A	exon3
Chr08	*Glyma.08G299500*	41769114	A	C	exon3
Chr08	*Glyma.08G299500*	41769173	G	A	exon3
Chr08	*Glyma.08G299800*	41796279	T	C	exon2
Chr08	*Glyma.08G300200*	41828898	C	T	exon1
Chr08	*Glyma.08G300400*	41838405	C	T	exon4
Chr08	*Glyma.08G300500*	41866305	A	G	exon1
Chr08	*Glyma.08G301300*	41932498	T	C	exon3
Chr08	*Glyma.08G301700*	41965711	T	G	exon5
Chr08	*Glyma.08G301800*	41975248	G	C	exon6
Chr08	*Glyma.08G301800*	41975488	C	T	exon6
Chr08	*Glyma.08G301800*	41977127	C	G	exon2
Chr08	*Glyma.08G301800*	41977293	A	T	exon2

Chr08	*Glyma.08G301800*	41977547	A	G	exon2
Chr08	*Glyma.08G301800*	41977656	T	A	exon2
Chr08	*Glyma.08G301800*	41977836	C	A	exon1
Chr08	*Glyma.08G302400*	42047044	A	G	exon2
Chr08	*Glyma.08G302400*	42047237	A	G	exon2
Chr08	*Glyma.08G303000*	42116280	T	A	exon2
Chr08	*Glyma.08G303000*	42116295	T	G	exon2
Chr08	*Glyma.08G303000*	42116710	G	A	exon3
Chr08	*Glyma.08G303000*	42116911	C	G	exon3
Chr08	*Glyma.08G303500*	42150187	T	C	exon1
Chr08	*Glyma.08G303500*	42150987	T	G	exon1
Chr08	*Glyma.08G303500*	42150994	C	A	exon1
Chr08	*Glyma.08G303600*	42160203	A	G	exon5
Chr08	*Glyma.08G303600*	42160222	T	C	exon5
Chr08	*Glyma.08G303600*	42160312	A	G	exon5
Chr08	*Glyma.08G303600*	42161324	T	C	exon2
Chr08	*Glyma.08G305600*	42376026	C	T	exon1
Chr08	*Glyma.08G307300*	42553110	C	T	exon1
Chr08	*Glyma.08G307400*	42557005	T	C	exon2
Chr08	*Glyma.08G307400*	42557990	A	C	exon1
Chr08	*Glyma.08G307400*	42558235	G	T	exon1
Chr08	*Glyma.08G307700*	42604828	G	A	exon3
Chr08	*Glyma.08G309200*	42820284	A	T	exon1
Chr10	*Glyma.10G142500*	37698907	C	T	exon4
Chr13	*Glyma.13G145000*	25763283	A	G	exon1
Chr13	*Glyma.13G211700*	32542733	G	C	exon1
Chr13	*Glyma.13G211700*	32542851	A	C	exon1
Chr13	*Glyma.13G211900*	32559767	G	A	exon13
Chr13	*Glyma.13G211900*	32560874	G	T	exon13
Chr13	*Glyma.13G211900*	32566336	G	A	exon3

Supplementary Table 3. A list of unique Indel variants identified in 11 sucrose-related QTL in PI 506593 on Wm82.a2.v1 reference genome assembly.

Chr	Gene ID_Wm82.a2.v1	Type	Position Start	Position End	Reference	Variant	Exon
Chr04	*Glyma.04G203700*	frameshift deletion	47638896	47638897	AA	-	exon1
Chr04	*Glyma.04G204200*	stopgain	47662714	47662714	-[1]	A	exon1
Chr04	*Glyma.04G206500*	frameshift insertion	47904729	47904729	-	A	exon1
Chr05	*Glyma.05G156400*	frameshift deletion	34890616	34890648	AGAAGGATAGGGGGGA AAGGCATTTGATAATTT	-	exon6
Chr06	*Glyma.06G199500*	frameshift deletion	18213487	18213494	AACTTCAC	-	exon4
Chr06	*Glyma.06G200200*	frameshift deletion	18329553	18329553	A	-	exon4
Chr06	*Glyma.06G200300*	frameshift insertion	18337703	18337703	-	C	exon2
Chr06	*Glyma.06G204000*	frameshift deletion	19151477	19151477	C	-	exon2
Chr06	*Glyma.06G205300*	frameshift deletion	19396876	19396876	C	-	exon4
Chr06	*Glyma.06G245900*	frameshift insertion	41535963	41535963	-	ACCGATAGAAACAA GTACATTGGTTAGC	exon2
Chr06	*Glyma.06G258900*	frameshift insertion	44382422	44382422	-	A	exon1
Chr06	*Glyma.06G258900*	frameshift insertion	44382444	44382444	-	GCCG	exon1
Chr06	*Glyma.06G259800*	nonframeshift deletion	44468739	44468756	GAGTCTTCTTCTAAGTCA	-	exon3
Chr06	*Glyma.06G259800*	nonframeshift deletion	44468799	44468801	TCT	-	exon3
Chr06	*Glyma.06G260100*	nonframeshift deletion	44524823	44524825	AGA	-	exon1
Chr06	*Glyma.06G267500*	frameshift insertion	45611906	45611906	-	C	exon3
Chr06	*Glyma.06G270600*	nonframeshift deletion	45982937	45982948	CTGTTCAACCAT	-	exon2
Chr06	*Glyma.06G276000*	nonframeshift deletion	46703961	46703969	GGTGCGGGT	-	exon1
Chr06	*Glyma.06G276400*	nonframeshift deletion	46764582	46764584	CAC	-	exon5
Chr06	*Glyma.06G276700*	nonframeshift insertion	46778678	46778678	-	TCCTCCACCGCCACATCC	exon1

Chr08	*Glyma.08G292200*	frameshift deletion	40620139	40620149	GCTTTTCCTTT	-	exon3
Chr08	*Glyma.08G293800*	nonframeshift deletion	40884492	40884509	CATAGTTAAGCCTATTCC	-	exon2
Chr08	*Glyma.08G295400*	nonframeshift deletion	41033092	41033094	TTT	-	exon1
Chr08	*Glyma.08G299800*	frameshift insertion	41796428	41796428	-	CCAAGTTCCTTAAATG	exon2
Chr08	*Glyma.08G300600*	nonframeshift insertion	41874915	41874915	-	AAT	exon1
Chr08	*Glyma.08G301700*	frameshift deletion	41966266	41966266	G	-	exon6
Chr08	*Glyma.08G301800*	frameshift deletion	41977619	41977619	T	-	exon2
Chr08	*Glyma.08G303000*	frameshift deletion	42116353	42116353	T	-	exon2
Chr08	*Glyma.08G303300*	frameshift deletion	42138703	42138704	CT	-	exon5
Chr08	*Glyma.08G303500*	nonframeshift deletion	42151425	42151445	TGAGATGATTGTCCAGATGAG	-	exon1
Chr08	*Glyma.08G304300*	frameshift deletion	42272971	42272971	C	-	exon4

[1]Data not available

Supplementary Table 4. A list of unique nonsynonymous SNP variants identified in 11 sucrose-related QTL in PI 506593 on Wm82.a5.v1 reference genome assembly and the corresponding gene ID on Wm82.a2.v1 reference genome assembly.

Chr	Gene ID_Wm82.a5.v1	Position	Reference	Variant	Exon	Gene_Wm82.a2.v1
Chr04	*Gm_Wm82_10951*	50012775	T	G	exon2	*Glyma.04g196400*
Chr04	*Gm_Wm82_10995*	50609653	A	G	exon3	*Glyma.04g201400*
Chr04	*Gm_Wm82_10997*	50630113	A	G	exon12	*Glyma.04g201600*
Chr04	*Gm_Wm82_10997*	50630118	A	G	exon12	*Glyma.04g201600*
Chr04	*Gm_Wm82_10997*	50632119	T	A	exon7	*Glyma.04g201600*
Chr04	*Gm_Wm82_10997*	50635152	A	G	exon2	*Glyma.04g201600*
Chr04	*Gm_Wm82_10997*	50635401	T	C	exon1	*Glyma.04g201600*
Chr04	*Gm_Wm82_10999*	50641216	T	C	exon1	-[1]
Chr04	*Gm_Wm82_10999*	50641591	A	G	exon2	-
Chr04	*Gm_Wm82_11000*	50643932	G	C	exon2	*Glyma.04g201800*
Chr04	*Gm_Wm82_11000*	50644042	T	G	exon2	*Glyma.04g201800*
Chr04	*Gm_Wm82_11007*	50703497	C	A	exon1	*Glyma.04g202600*
Chr04	*Gm_Wm82_11016*	50788943	C	T	exon5	*Glyma.04g203400*
Chr04	*Gm_Wm82_11018*	50807177	G	A	exon3	*Glyma.04g203600*
Chr04	*Gm_Wm82_11023*	50841137	T	G	exon6	*Glyma.04g204400*
Chr04	*Gm_Wm82_11028*	50918307	A	C	exon1	-
Chr04	*Gm_Wm82_11028*	50918331	C	A	exon1	-
Chr04	*Gm_Wm82_11028*	50918509	C	G	exon2	-
Chr04	*Gm_Wm82_11028*	50918576	T	C	exon2	-
Chr04	*Gm_Wm82_11029*	50930826	G	C	exon3	*Glyma.04g205000*
Chr04	*Gm_Wm82_11029*	50930935	G	A	exon3	*Glyma.04g205000*
Chr04	*Gm_Wm82_11029*	50930962	G	A	exon3	*Glyma.04g205000*
Chr04	*Gm_Wm82_11029*	50931456	A	G	exon3	*Glyma.04g205000*
Chr04	*Gm_Wm82_11029*	50931483	C	T	exon3	*Glyma.04g205000*
Chr04	*Gm_Wm82_11033*	50986583	T	C	exon1	*Glyma.04g205500*
Chr04	*Gm_Wm82_11038*	51032954	G	T	exon4	*Glyma.04g206000*
Chr04	*Gm_Wm82_11039*	51046670	A	T	exon6	*Glyma.04g206100*
Chr04	*Gm_Wm82_11041*	51066509	T	G	exon12	*Glyma.04g206400*
Chr04	*Gm_Wm82_11041*	51070267	C	T	exon8	*Glyma.04g206400*
Chr04	*Gm_Wm82_11041*	51070928	C	A	exon6	*Glyma.04g206400*
Chr04	*Gm_Wm82_11041*	51070938	C	T	exon6	*Glyma.04g206400*
Chr04	*Gm_Wm82_11041*	51070955	A	G	exon6	*Glyma.04g206400*
Chr04	*Gm_Wm82_11041*	51071938	G	A	exon5	*Glyma.04g206400*
Chr04	*Gm_Wm82_11041*	51073044	A	T	exon4	*Glyma.04g206400*
Chr04	*Gm_Wm82_11041*	51073542	A	G	exon2	*Glyma.04g206400*
Chr04	*Gm_Wm82_11041*	51074673	T	C	exon1	*Glyma.04g206400*
Chr04	*Gm_Wm82_11042*	51086506	T	A	exon1	-
Chr04	*Gm_Wm82_11043*	51097351	T	G	exon1	*Glyma.04g206700*
Chr04	*Gm_Wm82_11043*	51097989	G	A	exon1	*Glyma.04g206700*
Chr04	*Gm_Wm82_11045*	51127707	C	A	exon1	*Glyma.04g206900*
Chr04	*Gm_Wm82_11045*	51127712	T	C	exon1	*Glyma.04g206900*
Chr05	*Gm_Wm82_13207*	36648649	G	A	exon3	-
Chr05	*Gm_Wm82_13210*	36656386	A	T	exon5	*Glyma.05g149500*
Chr05	*Gm_Wm82_13210*	36656901	T	C	exon4	*Glyma.05g149500*
Chr05	*Gm_Wm82_13210*	36657267	G	C	exon3	*Glyma.05g149500*
Chr05	*Gm_Wm82_13210*	36658697	C	A	exon1	*Glyma.05g149500*
Chr05	*Gm_Wm82_13210*	36658702	C	T	exon1	*Glyma.05g149500*
Chr05	*Gm_Wm82_13210*	36658734	G	A	exon1	*Glyma.05g149500*
Chr05	*Gm_Wm82_13211*	36660650	G	A	exon2	*Glyma.05g149600*
Chr05	*Gm_Wm82_13211*	36660945	A	G	exon2	*Glyma.05g149600*
Chr05	*Gm_Wm82_13212*	36664436	T	C	exon12	*Glyma.05g149700*
Chr05	*Gm_Wm82_13212*	36669292	G	A	exon1	*Glyma.05g149700*
Chr05	*Gm_Wm82_13224*	36744603	T	A	exon1	*Glyma.05g150800*
Chr05	*Gm_Wm82_13248*	36977483	G	A	exon2	*Glyma.05g153200*
Chr05	*Gm_Wm82_13251*	36998204	G	T	exon2	*Glyma.05g153700*
Chr05	*Gm_Wm82_13271*	37153828	A	G	exon4	*Glyma.05g156000*

Chr05	*Gm_Wm82_13271*	37155841	A	T	exon2	*Glyma.05g156000*
Chr05	*Gm_Wm82_13275*	37174510	A	C	exon5	*Glyma.05g156400*
Chr05	*Gm_Wm82_13276*	37179047	A	C	exon1	*Glyma.05g156500*
Chr05	*Gm_Wm82_13276*	37179088	T	A	exon1	*Glyma.05g156500*
Chr05	*Gm_Wm82_13276*	37179340	C	T	exon1	*Glyma.05g156500*
Chr05	*Gm_Wm82_13276*	37179395	T	A	exon1	*Glyma.05g156500*
Chr05	*Gm_Wm82_13276*	37179689	C	G	exon1	*Glyma.05g156500*
Chr05	*Gm_Wm82_13277*	37198511	A	G	exon1	*Glyma.05g156700*
Chr05	*Gm_Wm82_13287*	37304314	G	A	exon4	*Glyma.05g158000*
Chr06	*Gm_Wm82_15854*	16855065	T	A	exon10	*Glyma.06g191300*
Chr06	*Gm_Wm82_15925*	17691298	A	G	exon5	-
Chr06	*Gm_Wm82_15925*	17692083	A	C	exon2	-
Chr06	*Gm_Wm82_15950*	18125603	C	T	exon9	*Glyma.06g199400*
Chr06	*Gm_Wm82_15950*	18125738	G	C	exon8	*Glyma.06g199400*
Chr06	*Gm_Wm82_15950*	18126225	T	A	exon6	*Glyma.06g199400*
Chr06	*Gm_Wm82_15958*	18318253	C	T	exon4	*Glyma.06g200800*
Chr06	*Gm_Wm82_15959*	18332225	A	C	exon1	*Glyma.06g201000*
Chr06	*Gm_Wm82_15995*	19190984	C	G	exon3	*Glyma.06g204400*
Chr06	*Gm_Wm82_15997*	19251705	A	G	exon1	*Glyma.06g204600*
Chr06	*Gm_Wm82_15997*	19252121	G	A	exon1	*Glyma.06g204600*
Chr06	*Gm_Wm82_16000*	19291120	A	G	exon1	*Glyma.06g204900*
Chr06	*Gm_Wm82_16000*	19291174	A	G	exon1	*Glyma.06g204900*
Chr06	*Gm_Wm82_16000*	19291274	T	A	exon1	*Glyma.06g204900*
Chr06	*Gm_Wm82_16000*	19291282	T	C	exon1	*Glyma.06g204900*
Chr06	*Gm_Wm82_16000*	19291307	G	A	exon1	*Glyma.06g204900*
Chr06	*Gm_Wm82_16000*	19291313	T	C	exon1	*Glyma.06g204900*
Chr06	*Gm_Wm82_16000*	19291316	G	T	exon1	*Glyma.06g204900*
Chr06	*Gm_Wm82_16007*	19398634	T	C	exon2	*Glyma.06g205600*
Chr06	*Gm_Wm82_16028*	19890710	T	G	exon9	*Glyma.06g207000*
Chr06	*Gm_Wm82_16028*	19890712	C	T	exon9	*Glyma.06g207000*
Chr06	*Gm_Wm82_16028*	19898094	G	A	exon13	*Glyma.06g207000*
Chr06	*Gm_Wm82_16031*	19949286	G	A	exon7	*Glyma.06g207300*
Chr06	*Gm_Wm82_16037*	20060581	C	A	exon2	-
Chr06	*Gm_Wm82_16037*	20060802	A	G	exon2	-
Chr06	*Gm_Wm82_16037*	20060821	T	A	exon2	-
Chr06	*Gm_Wm82_16037*	20060836	T	C	exon2	-
Chr06	*Gm_Wm82_16037*	20060844	A	G	exon2	-
Chr06	*Gm_Wm82_16037*	20060853	T	G	exon2	-
Chr06	*Gm_Wm82_16037*	20060857	A	G	exon2	-
Chr06	*Gm_Wm82_16037*	20060875	T	C	exon2	-
Chr06	*Gm_Wm82_16037*	20060880	A	G	exon2	-
Chr06	*Gm_Wm82_16037*	20060885	A	C	exon2	-
Chr06	*Gm_Wm82_16093*	21318042	A	G	exon1	*Glyma.06g213600*
Chr06	*Gm_Wm82_16114*	21672346	G	A	exon5	-
Chr06	*Gm_Wm82_16114*	21672369	A	T	exon5	-
Chr06	*Gm_Wm82_16124*	21964101	A	G	exon6	*Glyma.06g214500*
Chr06	*Gm_Wm82_16124*	21964111	A	C	exon6	*Glyma.06g214500*
Chr06	*Gm_Wm82_16124*	21964210	T	C	exon6	*Glyma.06g214500*
Chr06	*Gm_Wm82_16124*	21964256	T	A	exon6	*Glyma.06g214500*
Chr06	*Gm_Wm82_16124*	21964371	A	G	exon5	*Glyma.06g214500*
Chr06	*Gm_Wm82_16130*	22146914	G	A	exon11	*Glyma.06g215100*
Chr06	*Gm_Wm82_16130*	22150066	T	G	exon4	*Glyma.06g215100*
Chr06	*Gm_Wm82_16134*	22242342	G	A	exon7	*Glyma.06g215300*
Chr06	*Gm_Wm82_16134*	22245731	G	A	exon1	*Glyma.06g215300*
Chr06	*Gm_Wm82_16134*	22245752	T	A	exon1	*Glyma.06g215300*
Chr06	*Gm_Wm82_16135*	22247215	T	C	exon3	*Glyma.06g215300*
Chr06	*Gm_Wm82_16144*	22449065	G	A	exon2	*Glyma.06g216100*
Chr06	*Gm_Wm82_16163*	22903124	G	A	exon2	*Glyma.06g217400*
Chr06	*Gm_Wm82_16190*	24943595	A	G	exon3	*Glyma.06g220600*
Chr06	*Gm_Wm82_16206*	27167620	C	G	exon1	*Glyma.06g221400*
Chr06	*Gm_Wm82_16222*	28810357	G	T	exon1	-
Chr06	*Gm_Wm82_16237*	29130322	C	G	exon1	-

Chr06	*Gm_Wm82_16239*	29191758	G	A	exon15	*Glyma.06g223100*
Chr06	*Gm_Wm82_16260*	30393588	C	T	exon1	*Glyma.06g224800*
Chr06	*Gm_Wm82_16275*	34385114	C	T	exon1	*Glyma.06g226400*
Chr06	*Gm_Wm82_16321*	36946425	C	T	exon4	-
Chr06	*Gm_Wm82_16321*	36946445	A	G	exon4	-
Chr06	*Gm_Wm82_16321*	36946458	T	G	exon4	-
Chr06	*Gm_Wm82_16369*	38608306	C	G	exon1	-
Chr06	*Gm_Wm82_16369*	38608336	A	G	exon1	-
Chr06	*Gm_Wm82_16369*	38608355	C	T	exon1	-
Chr06	*Gm_Wm82_16372*	38641979	C	A	exon3	*Glyma.06g233100*
Chr06	*Gm_Wm82_16418*	40254102	T	C	exon2	-
Chr06	*Gm_Wm82_16419*	40371913	A	T	exon10	-
Chr06	*Gm_Wm82_16427*	40473003	C	T	exon1	-
Chr06	*Gm_Wm82_16427*	40473051	C	T	exon1	-
Chr06	*Gm_Wm82_16446*	40912651	C	T	exon8	*Glyma.06g239000*
Chr06	*Gm_Wm82_16446*	40912654	G	A	exon8	*Glyma.06g239000*
Chr06	*Gm_Wm82_16446*	40912674	A	C	exon8	*Glyma.06g239000*
Chr06	*Gm_Wm82_16473*	41570644	A	C	exon5	*Glyma.06g241200*
Chr06	*Gm_Wm82_16525*	42513821	G	T	exon1	*Glyma.06g244200*
Chr06	*Gm_Wm82_16533*	42996668	G	A	exon1	-
Chr06	*Gm_Wm82_16540*	43189448	A	G	exon2	-
Chr06	*Gm_Wm82_16552*	43249640	C	A	exon2	-
Chr06	*Gm_Wm82_16552*	43249754	G	A	exon2	-
Chr06	*Gm_Wm82_16553*	43255775	T	C	exon2	*Glyma.06g245700*
Chr06	*Gm_Wm82_16712*	45704770	C	T	exon5	*Glyma.06g256000*
Chr06	*Gm_Wm82_16712*	45704787	C	T	exon5	*Glyma.06g256000*
Chr06	*Gm_Wm82_16712*	45704820	T	G	exon5	*Glyma.06g256000*
Chr06	*Gm_Wm82_16712*	45704842	C	A	exon5	*Glyma.06g256000*
Chr06	*Gm_Wm82_16712*	45704847	G	T	exon5	*Glyma.06g256000*
Chr06	*Gm_Wm82_16712*	45704849	A	T	exon5	*Glyma.06g256000*
Chr06	*Gm_Wm82_16712*	45704859	T	C	exon5	*Glyma.06g256000*
Chr06	*Gm_Wm82_16712*	45704865	T	A	exon5	*Glyma.06g256000*
Chr06	*Gm_Wm82_16712*	45705132	C	A	exon5	*Glyma.06g256000*
Chr06	*Gm_Wm82_16712*	45705138	G	A	exon5	*Glyma.06g256000*
Chr06	*Gm_Wm82_16712*	45705163	C	T	exon5	*Glyma.06g256000*
Chr06	*Gm_Wm82_16712*	45705167	C	T	exon5	*Glyma.06g256000*
Chr06	*Gm_Wm82_16712*	45705186	C	G	exon5	*Glyma.06g256000*
Chr06	*Gm_Wm82_16712*	45705199	G	C	exon5	*Glyma.06g256000*
Chr06	*Gm_Wm82_16712*	45705201	T	C	exon5	*Glyma.06g256000*
Chr06	*Gm_Wm82_16712*	45705265	T	A	exon5	*Glyma.06g256000*
Chr06	*Gm_Wm82_16712*	45706662	C	T	exon3	*Glyma.06g256000*
Chr06	*Gm_Wm82_16712*	45706674	C	T	exon3	*Glyma.06g256000*
Chr06	*Gm_Wm82_16712*	45706676	T	C	exon3	*Glyma.06g256000*
Chr06	*Gm_Wm82_16712*	45706683	G	T	exon3	*Glyma.06g256000*
Chr06	*Gm_Wm82_16712*	45706688	A	G	exon3	*Glyma.06g256000*
Chr06	*Gm_Wm82_16712*	45706692	G	T	exon3	*Glyma.06g256000*
Chr06	*Gm_Wm82_16712*	45706695	C	T	exon3	*Glyma.06g256000*
Chr06	*Gm_Wm82_16712*	45706703	G	A	exon3	*Glyma.06g256000*
Chr06	*Gm_Wm82_16712*	45706719	C	T	exon3	*Glyma.06g256000*
Chr06	*Gm_Wm82_16712*	45706722	G	C	exon3	*Glyma.06g256000*
Chr06	*Gm_Wm82_16712*	45708230	G	A	exon2	*Glyma.06g256000*
Chr06	*Gm_Wm82_16712*	45708317	G	A	exon2	*Glyma.06g256000*
Chr06	*Gm_Wm82_16712*	45708359	A	T	exon2	*Glyma.06g256000*
Chr06	*Gm_Wm82_16714*	45717132	G	T	exon4	*Glyma.06g256200*
Chr06	*Gm_Wm82_16714*	45717139	C	A	exon4	*Glyma.06g256200*
Chr06	*Gm_Wm82_16714*	45717146	G	T	exon4	*Glyma.06g256200*
Chr06	*Gm_Wm82_16714*	45717159	T	A	exon4	*Glyma.06g256200*
Chr06	*Gm_Wm82_16714*	45717168	G	T	exon4	*Glyma.06g256200*
Chr06	*Gm_Wm82_16714*	45717229	G	A	exon4	*Glyma.06g256200*
Chr06	*Gm_Wm82_16714*	45717235	G	T	exon4	*Glyma.06g256200*
Chr06	*Gm_Wm82_16714*	45717249	T	G	exon4	*Glyma.06g256200*
Chr06	*Gm_Wm82_16715*	45722376	G	A	exon3	-

Chr06	*Gm_Wm82_16721*	45802251	A	T	exon2	*Glyma.06g256600*
Chr06	*Gm_Wm82_16723*	45828405	T	G	exon1	*Glyma.06g256700*
Chr06	*Gm_Wm82_16723*	45833001	A	T	exon5	*Glyma.06g256700*
Chr06	*Gm_Wm82_16737*	46055491	C	T	exon1	-
Chr06	*Gm_Wm82_16741*	46120780	C	T	exon2	-
Chr06	*Gm_Wm82_16770*	46519180	G	T	exon8	*Glyma.06g258800*
Chr06	*Gm_Wm82_16770*	46523865	G	A	exon1	*Glyma.06g258800*
Chr06	*Gm_Wm82_16771*	46528082	G	T	exon1	*Glyma.06g258900*
Chr06	*Gm_Wm82_16771*	46528122	C	T	exon1	*Glyma.06g258900*
Chr06	*Gm_Wm82_16771*	46528174	A	T	exon1	*Glyma.06g258900*
Chr06	*Gm_Wm82_16771*	46528179	A	G	exon1	*Glyma.06g258900*
Chr06	*Gm_Wm82_16771*	46528253	G	C	exon1	*Glyma.06g258900*
Chr06	*Gm_Wm82_16771*	46528256	A	G	exon1	*Glyma.06g258900*
Chr06	*Gm_Wm82_16771*	46528260	T	C	exon1	*Glyma.06g258900*
Chr06	*Gm_Wm82_16771*	46528340	C	A	exon1	*Glyma.06g258900*
Chr06	*Gm_Wm82_16771*	46528441	T	A	exon1	*Glyma.06g258900*
Chr06	*Gm_Wm82_16771*	46528466	G	T	exon1	*Glyma.06g258900*
Chr06	*Gm_Wm82_16771*	46528557	A	T	exon2	*Glyma.06g258900*
Chr06	*Gm_Wm82_16771*	46528575	T	C	exon2	*Glyma.06g258900*
Chr06	*Gm_Wm82_16771*	46528605	G	A	exon2	*Glyma.06g258900*
Chr06	*Gm_Wm82_16771*	46529547	A	T	exon3	*Glyma.06g258900*
Chr06	*Gm_Wm82_16771*	46529627	G	A	exon3	*Glyma.06g258900*
Chr06	*Gm_Wm82_16771*	46529825	G	T	exon4	*Glyma.06g258900*
Chr06	*Gm_Wm82_16771*	46529831	G	A	exon4	*Glyma.06g258900*
Chr06	*Gm_Wm82_16771*	46529906	A	G	exon4	*Glyma.06g258900*
Chr06	*Gm_Wm82_16771*	46529927	A	C	exon4	*Glyma.06g258900*
Chr06	*Gm_Wm82_16771*	46530174	T	C	exon5	*Glyma.06g258900*
Chr06	*Gm_Wm82_16771*	46530265	G	T	exon5	*Glyma.06g258900*
Chr06	*Gm_Wm82_16771*	46530485	T	A	exon6	*Glyma.06g258900*
Chr06	*Gm_Wm82_16771*	46530569	G	A	exon6	*Glyma.06g258900*
Chr06	*Gm_Wm82_16771*	46530585	G	A	exon6	*Glyma.06g258900*
Chr06	*Gm_Wm82_16771*	46531051	G	A	exon8	*Glyma.06g258900*
Chr06	*Gm_Wm82_16771*	46531176	T	A	exon8	*Glyma.06g258900*
Chr06	*Gm_Wm82_16771*	46531228	A	C	exon8	*Glyma.06g258900*
Chr06	*Gm_Wm82_16773*	46559716	G	C	exon5	-
Chr06	*Gm_Wm82_16773*	46559728	T	C	exon5	-
Chr06	*Gm_Wm82_16773*	46559747	G	T	exon5	-
Chr06	*Gm_Wm82_16773*	46559767	G	C	exon5	-
Chr06	*Gm_Wm82_16773*	46559798	C	G	exon5	-
Chr06	*Gm_Wm82_16773*	46559932	C	T	exon4	-
Chr06	*Gm_Wm82_16773*	46559963	T	C	exon4	-
Chr06	*Gm_Wm82_16773*	46559995	A	C	exon4	-
Chr06	*Gm_Wm82_16773*	46560011	C	T	exon4	-
Chr06	*Gm_Wm82_16773*	46560021	G	A	exon4	-
Chr06	*Gm_Wm82_16773*	46560029	G	C	exon4	-
Chr06	*Gm_Wm82_16773*	46560730	C	T	exon4	-
Chr06	*Gm_Wm82_16773*	46587448	A	G	exon2	-
Chr06	*Gm_Wm82_16773*	46675331	C	T	exon1	-
Chr06	*Gm_Wm82_16773*	46675332	T	C	exon1	-
Chr06	*Gm_Wm82_16773*	46675428	A	C	exon1	-
Chr06	*Gm_Wm82_16773*	46675440	C	G	exon1	-
Chr06	*Gm_Wm82_16773*	46675509	A	G	exon1	-
Chr06	*Gm_Wm82_16773*	46675517	C	T	exon1	-
Chr06	*Gm_Wm82_16773*	46675593	G	C	exon1	-
Chr06	*Gm_Wm82_16773*	46675770	C	G	exon1	-
Chr06	*Gm_Wm82_16781*	46816050	A	G	exon1	*Glyma.06g260900*
Chr06	*Gm_Wm82_16781*	46816321	C	T	exon1	*Glyma.06g260900*
Chr06	*Gm_Wm82_16781*	46816376	C	A	exon1	*Glyma.06g260900*
Chr06	*Gm_Wm82_16783*	46849360	G	A	exon2	-
Chr06	*Gm_Wm82_16784*	46875167	G	A	exon1	*Glyma.06g261100*
Chr06	*Gm_Wm82_16784*	46875224	T	A	exon1	*Glyma.06g261100*
Chr06	*Gm_Wm82_16784*	46875335	C	T	exon1	*Glyma.06g261100*

Chr06	*Gm_Wm82_16788*	46938940	T	C	exon2	*Glyma.06g261400*
Chr06	*Gm_Wm82_16788*	46939025	G	A	exon2	*Glyma.06g261400*
Chr06	*Gm_Wm82_16788*	46939148	A	G	exon2	*Glyma.06g261400*
Chr06	*Gm_Wm82_16788*	46939679	G	T	exon2	*Glyma.06g261400*
Chr06	*Gm_Wm82_16788*	46939724	G	A	exon2	*Glyma.06g261400*
Chr06	*Gm_Wm82_16788*	46939728	A	T	exon2	*Glyma.06g261400*
Chr06	*Gm_Wm82_16788*	46939733	C	G	exon2	*Glyma.06g261400*
Chr06	*Gm_Wm82_16788*	46939745	T	A	exon2	*Glyma.06g261400*
Chr06	*Gm_Wm82_16788*	46939778	G	T	exon2	*Glyma.06g261400*
Chr06	*Gm_Wm82_16788*	46980614	T	A	exon1	*Glyma.06g261400*
Chr06	*Gm_Wm82_16788*	46980694	G	A	exon1	*Glyma.06g261400*
Chr06	*Gm_Wm82_16788*	46980719	C	A	exon1	*Glyma.06g261400*
Chr06	*Gm_Wm82_16788*	46980727	T	A	exon1	*Glyma.06g261400*
Chr06	*Gm_Wm82_16788*	46980741	A	T	exon1	*Glyma.06g261400*
Chr06	*Gm_Wm82_16788*	46980793	T	C	exon1	*Glyma.06g261400*
Chr06	*Gm_Wm82_16788*	46980812	C	T	exon1	*Glyma.06g261400*
Chr06	*Gm_Wm82_16788*	46981016	G	T	exon1	*Glyma.06g261400*
Chr06	*Gm_Wm82_16788*	46981024	T	C	exon1	*Glyma.06g261400*
Chr06	*Gm_Wm82_16788*	46981030	C	T	exon1	*Glyma.06g261400*
Chr06	*Gm_Wm82_16792*	47017537	A	G	exon8	*Glyma.06g261800*
Chr06	*Gm_Wm82_16810*	47272369	T	C	exon1	*Glyma.06g263800*
Chr06	*Gm_Wm82_16810*	47272426	C	T	exon1	*Glyma.06g263800*
Chr06	*Gm_Wm82_16810*	47272441	T	C	exon1	*Glyma.06g263800*
Chr06	*Gm_Wm82_16810*	47272525	T	C	exon1	*Glyma.06g263800*
Chr06	*Gm_Wm82_16810*	47272534	A	T	exon1	*Glyma.06g263800*
Chr06	*Gm_Wm82_16810*	47272549	T	C	exon1	*Glyma.06g263800*
Chr06	*Gm_Wm82_16810*	47273831	G	A	exon4	*Glyma.06g263800*
Chr06	*Gm_Wm82_16810*	47273837	T	C	exon4	*Glyma.06g263800*
Chr06	*Gm_Wm82_16810*	47274025	C	T	exon5	*Glyma.06g263800*
Chr06	*Gm_Wm82_16810*	47274662	C	T	exon7	*Glyma.06g263800*
Chr06	*Gm_Wm82_16810*	47274951	C	A	exon8	*Glyma.06g263800*
Chr06	*Gm_Wm82_16810*	47274989	A	G	exon8	*Glyma.06g263800*
Chr06	*Gm_Wm82_16818*	47435150	G	A	exon2	*Glyma.06g265400*
Chr06	*Gm_Wm82_16818*	47435168	A	G	exon2	*Glyma.06g265400*
Chr06	*Gm_Wm82_16818*	47435180	A	C	exon2	*Glyma.06g265400*
Chr06	*Gm_Wm82_16818*	47435193	G	A	exon2	*Glyma.06g265400*
Chr06	*Gm_Wm82_16819*	47437782	C	G	exon1	*Glyma.06g265400*
Chr06	*Gm_Wm82_16819*	47437788	T	G	exon1	*Glyma.06g265400*
Chr06	*Gm_Wm82_16819*	47437797	G	A	exon1	*Glyma.06g265400*
Chr06	*Gm_Wm82_16819*	47437819	T	G	exon1	*Glyma.06g265400*
Chr06	*Gm_Wm82_16819*	47437845	T	A	exon1	*Glyma.06g265400*
Chr06	*Gm_Wm82_16819*	47437899	A	G	exon1	*Glyma.06g265400*
Chr06	*Gm_Wm82_16819*	47437912	A	G	exon1	*Glyma.06g265400*
Chr06	*Gm_Wm82_16819*	47438062	T	A	exon1	*Glyma.06g265400*
Chr06	*Gm_Wm82_16819*	47438064	T	A	exon1	*Glyma.06g265400*
Chr06	*Gm_Wm82_16819*	47438070	C	T	exon1	*Glyma.06g265400*
Chr06	*Gm_Wm82_16819*	47438071	C	T	exon1	*Glyma.06g265400*
Chr06	*Gm_Wm82_16819*	47438082	G	T	exon1	*Glyma.06g265400*
Chr06	*Gm_Wm82_16819*	47438085	T	G	exon1	*Glyma.06g265400*
Chr06	*Gm_Wm82_16823*	47492256	G	T	exon1	*Glyma.06g265500*
Chr06	*Gm_Wm82_16832*	47592881	C	T	exon2	*Glyma.06g266300*
Chr06	*Gm_Wm82_16837*	47647081	T	C	exon6	*Glyma.06g267000*
Chr06	*Gm_Wm82_16837*	47647965	C	T	exon5	*Glyma.06g267000*
Chr06	*Gm_Wm82_16837*	47649457	C	T	exon2	*Glyma.06g267000*
Chr06	*Gm_Wm82_16837*	47649580	T	C	exon2	*Glyma.06g267000*
Chr06	*Gm_Wm82_16843*	47792460	C	T	exon3	*Glyma.06g267900*
Chr06	*Gm_Wm82_16843*	47792464	G	A	exon3	*Glyma.06g267900*
Chr06	*Gm_Wm82_16845*	47845766	T	A	exon4	*Glyma.06g268700*
Chr06	*Gm_Wm82_16845*	47846225	G	A	exon4	*Glyma.06g268700*
Chr06	*Gm_Wm82_16845*	47846246	G	C	exon4	*Glyma.06g268700*
Chr06	*Gm_Wm82_16845*	47878412	A	T	exon3	*Glyma.06g268700*
Chr06	*Gm_Wm82_16845*	47878418	C	T	exon3	*Glyma.06g268700*

Chr06	*Gm_Wm82_16845*	47879216	A	C	exon2	*Glyma.06g268700*
Chr06	*Gm_Wm82_16845*	47879219	C	T	exon2	*Glyma.06g268700*
Chr06	*Gm_Wm82_16845*	47879230	A	G	exon2	*Glyma.06g268700*
Chr06	*Gm_Wm82_16845*	47879249	C	T	exon2	*Glyma.06g268700*
Chr06	*Gm_Wm82_16845*	47879259	G	C	exon2	*Glyma.06g268700*
Chr06	*Gm_Wm82_16845*	47879260	T	C	exon2	*Glyma.06g268700*
Chr06	*Gm_Wm82_16845*	47879285	C	T	exon2	*Glyma.06g268700*
Chr06	*Gm_Wm82_16845*	47879294	G	A	exon2	*Glyma.06g268700*
Chr06	*Gm_Wm82_16845*	47879296	T	C	exon2	*Glyma.06g268700*
Chr06	*Gm_Wm82_16847*	47948832	T	C	exon1	-
Chr06	*Gm_Wm82_16847*	47948901	A	G	exon1	-
Chr06	*Gm_Wm82_16847*	47948929	C	A	exon1	-
Chr06	*Gm_Wm82_16847*	47948931	A	T	exon1	-
Chr06	*Gm_Wm82_16847*	47948937	G	A	exon1	-
Chr06	*Gm_Wm82_16847*	47948947	A	G	exon1	-
Chr06	*Gm_Wm82_16847*	47948953	C	T	exon1	-
Chr06	*Gm_Wm82_16864*	48079838	G	A	exon1	-
Chr06	*Gm_Wm82_16864*	48079842	C	G	exon1	-
Chr06	*Gm_Wm82_16864*	48079874	G	A	exon1	-
Chr06	*Gm_Wm82_16864*	48079896	C	T	exon1	-
Chr06	*Gm_Wm82_16864*	48079901	G	C	exon1	-
Chr06	*Gm_Wm82_16864*	48080001	A	G	exon1	-
Chr06	*Gm_Wm82_16864*	48080057	G	A	exon1	-
Chr06	*Gm_Wm82_16865*	48110174	G	A	exon1	*Glyma.06g270400*
Chr06	*Gm_Wm82_16867*	48118112	A	G	exon1	-
Chr06	*Gm_Wm82_16867*	48118134	C	T	exon1	-
Chr06	*Gm_Wm82_16867*	48118148	A	G	exon1	-
Chr06	*Gm_Wm82_16867*	48118160	C	A	exon1	-
Chr06	*Gm_Wm82_16867*	48118169	A	G	exon1	-
Chr06	*Gm_Wm82_16867*	48118232	G	A	exon1	-
Chr06	*Gm_Wm82_16867*	48118235	G	A	exon1	-
Chr06	*Gm_Wm82_16867*	48118239	A	G	exon1	-
Chr06	*Gm_Wm82_16867*	48118253	G	A	exon1	-
Chr06	*Gm_Wm82_16871*	48154212	A	C	exon1	*Glyma.06g270900*
Chr06	*Gm_Wm82_16871*	48154217	G	A	exon1	*Glyma.06g270900*
Chr06	*Gm_Wm82_16871*	48154239	T	C	exon1	*Glyma.06g270900*
Chr06	*Gm_Wm82_16871*	48154245	T	C	exon1	*Glyma.06g270900*
Chr06	*Gm_Wm82_16871*	48154253	T	C	exon1	*Glyma.06g270900*
Chr06	*Gm_Wm82_16871*	48154320	T	C	exon1	*Glyma.06g270900*
Chr06	*Gm_Wm82_16871*	48154322	G	A	exon1	*Glyma.06g270900*
Chr06	*Gm_Wm82_16871*	48154362	C	T	exon1	*Glyma.06g270900*
Chr06	*Gm_Wm82_16871*	48159440	A	G	exon5	*Glyma.06g270900*
Chr06	*Gm_Wm82_16871*	48159446	A	T	exon5	*Glyma.06g270900*
Chr06	*Gm_Wm82_16871*	48159469	C	G	exon5	*Glyma.06g270900*
Chr06	*Gm_Wm82_16871*	48159476	T	C	exon5	*Glyma.06g270900*
Chr06	*Gm_Wm82_16871*	48159478	G	A	exon5	*Glyma.06g270900*
Chr06	*Gm_Wm82_16871*	48159511	A	G	exon5	*Glyma.06g270900*
Chr06	*Gm_Wm82_16871*	48159521	C	T	exon5	*Glyma.06g270900*
Chr06	*Gm_Wm82_16871*	48159529	A	T	exon5	*Glyma.06g270900*
Chr06	*Gm_Wm82_16872*	48164832	A	C	exon3	*Glyma.06g271000*
Chr06	*Gm_Wm82_16884*	48363825	G	A	exon2	*Glyma.06g272000*
Chr06	*Gm_Wm82_16886*	48431677	A	T	exon3	*Glyma.06g272200*
Chr06	*Gm_Wm82_16886*	48432051	T	A	exon3	*Glyma.06g272200*
Chr06	*Gm_Wm82_16887*	48444274	C	T	exon2	-
Chr06	*Gm_Wm82_16887*	48444309	G	A	exon2	-
Chr06	*Gm_Wm82_16887*	48444312	G	T	exon2	-
Chr06	*Gm_Wm82_16887*	48444439	G	A	exon1	-
Chr06	*Gm_Wm82_16887*	48444462	G	A	exon1	-
Chr06	*Gm_Wm82_16887*	48444495	C	T	exon1	-
Chr06	*Gm_Wm82_16887*	48444520	T	C	exon1	-
Chr06	*Gm_Wm82_16888*	48446801	A	G	exon1	-
Chr06	*Gm_Wm82_16888*	48446838	G	A	exon1	-

Chr06	*Gm_Wm82_16888*	48446925	T	C	exon1	-
Chr06	*Gm_Wm82_16888*	48446945	A	G	exon1	-
Chr06	*Gm_Wm82_16888*	48446972	T	C	exon1	-
Chr06	*Gm_Wm82_16888*	48446973	G	A	exon1	-
Chr06	*Gm_Wm82_16888*	48446976	G	A	exon1	-
Chr06	*Gm_Wm82_16888*	48447027	G	A	exon1	-
Chr06	*Gm_Wm82_16888*	48447101	C	G	exon1	-
Chr06	*Gm_Wm82_16889*	48447903	C	T	exon1	*Glyma.06g272300*
Chr06	*Gm_Wm82_16889*	48449996	A	G	exon4	*Glyma.06g272300*
Chr06	*Gm_Wm82_16889*	48450013	A	G	exon4	*Glyma.06g272300*
Chr06	*Gm_Wm82_16890*	48469114	G	A	exon3	*Glyma.06g272400*
Chr06	*Gm_Wm82_16906*	48610198	T	C	exon2	*Glyma.06g273700*
Chr06	*Gm_Wm82_16919*	48817377	A	C	exon12	*Glyma.06g275700*
Chr06	*Gm_Wm82_16919*	48817425	T	G	exon12	*Glyma.06g275700*
Chr06	*Gm_Wm82_16922*	48845516	A	G	exon1	*Glyma.06g276000*
Chr06	*Gm_Wm82_16922*	48845698	C	A	exon1	*Glyma.06g276000*
Chr06	*Gm_Wm82_16923*	48846930	A	G	exon2	-
Chr06	*Gm_Wm82_16923*	48847399	C	T	exon1	-
Chr06	*Gm_Wm82_16924*	48849091	T	G	exon5	*Glyma.06g276100*
Chr06	*Gm_Wm82_16924*	48849836	C	T	exon5	*Glyma.06g276100*
Chr06	*Gm_Wm82_16924*	48850120	T	C	exon5	*Glyma.06g276100*
Chr06	*Gm_Wm82_16929*	48939772	G	C	exon1	*Glyma.06g276800*
Chr06	*Gm_Wm82_16929*	48940064	C	G	exon1	*Glyma.06g276800*
Chr06	*Gm_Wm82_16929*	48941200	G	A	exon2	*Glyma.06g276800*
Chr08	*Gm_Wm82_23199*	43084507	G	A	exon3	-
Chr08	*Gm_Wm82_23200*	43094702	C	T	exon2	-
Chr08	*Gm_Wm82_23200*	43095133	T	G	exon4	-
Chr08	*Gm_Wm82_23200*	43095232	T	G	exon4	-
Chr08	*Gm_Wm82_23206*	43221710	C	T	exon14	*Glyma.08g293000*
Chr08	*Gm_Wm82_23206*	43224067	A	G	exon12	*Glyma.08g293000*
Chr08	*Gm_Wm82_23206*	43227388	C	T	exon4	*Glyma.08g293000*
Chr08	*Gm_Wm82_23209*	43253475	C	T	exon8	*Glyma.08g293100*
Chr08	*Gm_Wm82_23209*	43253484	G	A	exon8	*Glyma.08g293100*
Chr08	*Gm_Wm82_23209*	43259565	T	C	exon2	*Glyma.08g293100*
Chr08	*Gm_Wm82_23209*	43259653	C	T	exon2	*Glyma.08g293100*
Chr08	*Gm_Wm82_23209*	43259698	T	C	exon2	*Glyma.08g293100*
Chr08	*Gm_Wm82_23213*	43297301	A	T	exon3	-
Chr08	*Gm_Wm82_23213*	43297339	G	T	exon3	-
Chr08	*Gm_Wm82_23213*	43298027	T	G	exon1	-
Chr08	*Gm_Wm82_23216*	43328943	A	T	exon1	-
Chr08	*Gm_Wm82_23218*	43332689	C	T	exon4	-
Chr08	*Gm_Wm82_23220*	43345870	C	T	exon19	*Glyma.08g293900*
Chr08	*Gm_Wm82_23222*	43398382	A	T	exon3	*Glyma.08g294100*
Chr08	*Gm_Wm82_23222*	43399409	C	A	exon5	*Glyma.08g294100*
Chr08	*Gm_Wm82_23224*	43410052	C	A	exon5	*Glyma.08g294300*
Chr08	*Gm_Wm82_23226*	43442890	A	G	exon4	*Glyma.08g294700*
Chr08	*Gm_Wm82_23227*	43447814	A	G	exon1	*Glyma.08g294800*
Chr08	*Gm_Wm82_23229*	43465660	C	T	exon1	-
Chr08	*Gm_Wm82_23229*	43465661	G	T	exon1	-
Chr08	*Gm_Wm82_23230*	43479731	T	A	exon2	*Glyma.08g295400*
Chr08	*Gm_Wm82_23230*	43479791	A	T	exon2	*Glyma.08g295400*
Chr08	*Gm_Wm82_23231*	43486039	G	C	exon1	*Glyma.08g295400*
Chr08	*Gm_Wm82_23234*	43522304	C	G	exon5	*Glyma.08g295700*
Chr08	*Gm_Wm82_23235*	43529908	A	G	exon2	-
Chr08	*Gm_Wm82_23235*	43530059	A	G	exon3	-
Chr08	*Gm_Wm82_23235*	43530095	A	G	exon3	-
Chr08	*Gm_Wm82_23235*	43530107	A	G	exon3	-
Chr08	*Gm_Wm82_23235*	43530189	T	C	exon3	-
Chr08	*Gm_Wm82_23235*	43530607	T	C	exon4	-
Chr08	*Gm_Wm82_23235*	43530611	C	T	exon4	-
Chr08	*Gm_Wm82_23236*	43549493	T	A	exon2	-
Chr08	*Gm_Wm82_23236*	43549539	A	C	exon2	-

Chr08	*Gm_Wm82_23236*	43549677	A	G	exon2	-
Chr08	*Gm_Wm82_23237*	43554413	G	A	exon3	*Glyma.08g295800*
Chr08	*Gm_Wm82_23237*	43554470	T	C	exon3	*Glyma.08g295800*
Chr08	*Gm_Wm82_23243*	43670484	C	G	exon3	*Glyma.08g296300*
Chr08	*Gm_Wm82_23243*	43672465	A	G	exon5	*Glyma.08g296300*
Chr08	*Gm_Wm82_23245*	43694883	G	A	exon9	*Glyma.08g296500*
Chr08	*Gm_Wm82_23251*	43781623	A	C	exon8	*Glyma.08g297000*
Chr08	*Gm_Wm82_23255*	43813291	C	G	exon3	*Glyma.08g297400*
Chr08	*Gm_Wm82_23255*	43813989	C	T	exon1	*Glyma.08g297400*
Chr08	*Gm_Wm82_23256*	43832452	C	T	exon2	*Glyma.08g297500*
Chr08	*Gm_Wm82_23256*	43832655	C	G	exon2	*Glyma.08g297500*
Chr08	*Gm_Wm82_23257*	43861431	C	A	exon1	-
Chr08	*Gm_Wm82_23258*	43865006	T	A	exon1	*Glyma.08g297700*
Chr08	*Gm_Wm82_23258*	43865163	A	T	exon1	*Glyma.08g297700*
Chr08	*Gm_Wm82_23258*	43865196	A	G	exon1	*Glyma.08g297700*
Chr08	*Gm_Wm82_23258*	43865400	A	T	exon1	*Glyma.08g297700*
Chr08	*Gm_Wm82_23258*	43865468	G	C	exon1	*Glyma.08g297700*
Chr08	*Gm_Wm82_23258*	43868434	C	T	exon3	*Glyma.08g297700*
Chr08	*Gm_Wm82_23276*	44212054	T	C	exon3	*Glyma.08g299500*
Chr08	*Gm_Wm82_23276*	44212057	T	C	exon3	*Glyma.08g299500*
Chr08	*Gm_Wm82_23276*	44213496	G	A	exon2	*Glyma.08g299500*
Chr08	*Gm_Wm82_23276*	44213661	G	A	exon2	*Glyma.08g299500*
Chr08	*Gm_Wm82_23277*	44220882	C	T	exon3	*Glyma.08g299600*
Chr08	*Gm_Wm82_23278*	44228906	G	A	exon2	*Glyma.08g299700*
Chr08	*Gm_Wm82_23278*	44228949	A	G	exon2	*Glyma.08g299700*
Chr08	*Gm_Wm82_23279*	44240855	T	C	exon2	*Glyma.08g299800*
Chr08	*Gm_Wm82_23284*	44276755	T	C	exon3	*Glyma.08g300300*
Chr08	*Gm_Wm82_23285*	44282981	C	T	exon2	*Glyma.08g300400*
Chr08	*Gm_Wm82_23287*	44310884	A	G	exon1	*Glyma.08g300500*
Chr08	*Gm_Wm82_23293*	44376780	T	C	exon3	*Glyma.08g301300*
Chr08	*Gm_Wm82_23298*	44409909	T	G	exon5	-
Chr08	*Gm_Wm82_23299*	44444471	C	T	exon2	-
Chr08	*Gm_Wm82_23302*	44492044	A	G	exon1	*Glyma.08g302400*
Chr08	*Gm_Wm82_23302*	44492237	A	G	exon1	*Glyma.08g302400*
Chr08	*Gm_Wm82_23305*	44533472	T	C	exon5	*Glyma.08g302800*
Chr08	*Gm_Wm82_23307*	44559493	T	C	exon1	*Glyma.08g303000*
Chr08	*Gm_Wm82_23307*	44559499	C	T	exon1	*Glyma.08g303000*
Chr08	*Gm_Wm82_23307*	44559571	G	A	exon1	*Glyma.08g303000*
Chr08	*Gm_Wm82_23307*	44559572	G	A	exon1	*Glyma.08g303000*
Chr08	*Gm_Wm82_23307*	44559600	T	G	exon1	*Glyma.08g303000*
Chr08	*Gm_Wm82_23307*	44559604	G	A	exon1	*Glyma.08g303000*
Chr08	*Gm_Wm82_23307*	44559619	C	T	exon1	*Glyma.08g303000*
Chr08	*Gm_Wm82_23307*	44559622	G	A	exon1	*Glyma.08g303000*
Chr08	*Gm_Wm82_23307*	44559629	C	T	exon1	*Glyma.08g303000*
Chr08	*Gm_Wm82_23307*	44559644	C	T	exon1	*Glyma.08g303000*
Chr08	*Gm_Wm82_23307*	44559652	G	A	exon1	*Glyma.08g303000*
Chr08	*Gm_Wm82_23307*	44559721	G	A	exon1	*Glyma.08g303000*
Chr08	*Gm_Wm82_23307*	44559722	C	T	exon1	*Glyma.08g303000*
Chr08	*Gm_Wm82_23307*	44559730	G	A	exon1	*Glyma.08g303000*
Chr08	*Gm_Wm82_23307*	44559772	A	G	exon1	*Glyma.08g303000*
Chr08	*Gm_Wm82_23307*	44559776	C	T	exon1	*Glyma.08g303000*
Chr08	*Gm_Wm82_23307*	44561281	T	A	exon2	*Glyma.08g303000*
Chr08	*Gm_Wm82_23307*	44561296	T	G	exon2	*Glyma.08g303000*
Chr08	*Gm_Wm82_23307*	44561711	G	A	exon3	*Glyma.08g303000*
Chr08	*Gm_Wm82_23307*	44561912	C	G	exon3	*Glyma.08g303000*
Chr08	*Gm_Wm82_23312*	44595987	T	G	exon1	*Glyma.08g303500*
Chr08	*Gm_Wm82_23312*	44595994	C	A	exon1	*Glyma.08g303500*
Chr08	*Gm_Wm82_23312*	44596044	C	T	exon1	*Glyma.08g303500*
Chr08	*Gm_Wm82_23312*	44596051	T	A	exon1	*Glyma.08g303500*
Chr08	*Gm_Wm82_23312*	44596063	C	G	exon1	*Glyma.08g303500*
Chr08	*Gm_Wm82_23312*	44596170	T	C	exon1	*Glyma.08g303500*
Chr08	*Gm_Wm82_23312*	44596171	T	C	exon1	*Glyma.08g303500*

Chr08	*Gm_Wm82_23312*	44596179	G	A	exon1	*Glyma.08g303500*
Chr08	*Gm_Wm82_23312*	44596185	A	T	exon1	*Glyma.08g303500*
Chr08	*Gm_Wm82_23312*	44596191	A	T	exon1	*Glyma.08g303500*
Chr08	*Gm_Wm82_23312*	44596194	G	A	exon1	*Glyma.08g303500*
Chr08	*Gm_Wm82_23312*	44596206	T	A	exon1	*Glyma.08g303500*
Chr08	*Gm_Wm82_23312*	44596217	C	A	exon1	*Glyma.08g303500*
Chr08	*Gm_Wm82_23312*	44596234	C	A	exon1	*Glyma.08g303500*
Chr08	*Gm_Wm82_23312*	44596299	A	G	exon1	*Glyma.08g303500*
Chr08	*Gm_Wm82_23312*	44596314	C	T	exon1	*Glyma.08g303500*
Chr08	*Gm_Wm82_23312*	44596318	T	C	exon1	*Glyma.08g303500*
Chr08	*Gm_Wm82_23312*	44596320	T	C	exon1	*Glyma.08g303500*
Chr08	*Gm_Wm82_23312*	44596323	C	T	exon1	*Glyma.08g303500*
Chr08	*Gm_Wm82_23312*	44596334	C	A	exon1	*Glyma.08g303500*
Chr08	*Gm_Wm82_23312*	44596335	C	A	exon1	*Glyma.08g303500*
Chr08	*Gm_Wm82_23313*	44605203	A	G	exon8	*Glyma.08g303700*
Chr08	*Gm_Wm82_23313*	44605222	T	C	exon8	*Glyma.08g303700*
Chr08	*Gm_Wm82_23313*	44605312	A	G	exon8	*Glyma.08g303700*
Chr08	*Gm_Wm82_23313*	44606324	T	C	exon5	*Glyma.08g303700*
Chr08	*Gm_Wm82_23334*	44822836	C	T	exon1	*Glyma.08g305600*
Chr08	*Gm_Wm82_23345*	44921039	T	C	exon2	*Glyma.08g306700*
Chr08	*Gm_Wm82_23345*	44921087	G	A	exon2	*Glyma.08g306700*
Chr08	*Gm_Wm82_23351*	44991663	C	T	exon1	*Glyma.08g307300*
Chr08	*Gm_Wm82_23352*	44995560	T	C	exon2	*Glyma.08g307400*
Chr08	*Gm_Wm82_23352*	44995908	T	A	exon1	*Glyma.08g307400*
Chr08	*Gm_Wm82_23352*	44995941	C	T	exon1	*Glyma.08g307400*
Chr08	*Gm_Wm82_23352*	44995962	A	G	exon1	*Glyma.08g307400*
Chr08	*Gm_Wm82_23352*	44996545	A	C	exon1	*Glyma.08g307400*
Chr08	*Gm_Wm82_23352*	44996790	G	T	exon1	*Glyma.08g307400*
Chr08	*Gm_Wm82_23353*	45001133	A	C	exon1	-
Chr08	*Gm_Wm82_23356*	45029829	C	T	exon4	-
Chr08	*Gm_Wm82_23357*	45041930	G	A	exon3	*Glyma.08g307700*
Chr08	*Gm_Wm82_23372*	45257760	A	T	exon1	*Glyma.08g309200*
Chr10	*Gm_Wm82_28649*	41592195	C	T	exon3	*Glyma.10g142500*
Chr13	*Gm_Wm82_36817*	25667778	A	G	exon1	*Glyma.13g145000*
Chr13	*Gm_Wm82_37444*	32766267	G	A	exon13	*Glyma.13g211900*
Chr13	*Gm_Wm82_37444*	32772836	G	A	exon3	*Glyma.13g211900*

[1]Data not available

Supplementary Table 5. A list of unique Indel variants identified in 11 sucrose-related QTL in PI 506593 on Wm82.a5.v1 reference genome assembly and the corresponding gene ID on Wm82.a2.v1 reference genome assembly.

Chr	Gene ID Wm82.a5.v1	Type	Position Start	Position End	Reference	Variant	Exon	Gene ID Wm82.a2.v1
Chr04	*Gm_Wm82_10989*	nonframeshift deletion	50489090	50489092	ACC	-[1]	exon1	*Glyma.04g200500*
Chr04	*Gm_Wm82_11005*	nonframeshift deletion	50689703	50689705	AAG	-	exon3	*Glyma.04g202300*
Chr04	*Gm_Wm82_11016*	nonframeshift insertion	50790544	50790544	-	GTC	exon7	*Glyma.04g203400*
Chr04	*Gm_Wm82_11019*	frameshift deletion	50812339	50812340	AA	-	exon1	*Glyma.04g203700*
Chr04	*Gm_Wm82_11021*	stopgain	50836156	50836156	-	A	exon1	*Glyma.04g204200*
Chr04	*Gm_Wm82_11028*	frameshift insertion	50918389	50918389	-	C	exon1	-
Chr04	*Gm_Wm82_11041*	frameshift deletion	51070894	51070894	T	-	exon6	*Glyma.04g206400*
Chr04	*Gm_Wm82_11042*	frameshift deletion	51086279	51086279	G	-	exon1	-
Chr06	*Gm_Wm82_15925*	frameshift deletion	17691078	17691078	A	-	exon6	-
Chr06	*Gm_Wm82_15950*	frameshift deletion	18125755	18125756	TT	-	exon8	*Glyma.06g199400*
Chr06	*Gm_Wm82_15950*	frameshift deletion	18126193	18126200	CAACTTCA	-	exon6	*Glyma.06g199400*
Chr06	*Gm_Wm82_15991*	frameshift deletion	19089742	19089742	C	-	exon1	*Glyma.06g204000*
Chr06	*Gm_Wm82_15997*	nonframeshift insertion	19244539	19244539	-	TGA	exon8	*Glyma.06g204600*
Chr06	*Gm_Wm82_16037*	frameshift deletion	20060801	20060801	A	-	exon2	-
Chr06	*Gm_Wm82_16050*	frameshift insertion	20357956	20357956	-	AATT	exon2	-
Chr06	*Gm_Wm82_16072*	nonframeshift deletion	20782283	20782294	AACATCAACATC	-	exon1	-
Chr06	*Gm_Wm82_16321*	frameshift deletion	36946430	36946434	ACGTC	-	exon4	-
Chr06	*Gm_Wm82_16321*	nonframeshift deletion	36946469	36946474	GATGTT	-	exon4	-
Chr06	*Gm_Wm82_16322*	frameshift deletion	36951532	36951593	TTTGCTTCAAAAATC GGATCTTCGACGA TGGAATTAGCGC ACGAAGAAGAC	-	exon3	*Glyma.06g229800*

					AAACCCACCCG			
Chr06	*Gm_Wm82_16446*	frameshift deletion	40912657	40912661	TGTAG	-	exon8	*Glyma.06g239000*
Chr06	*Gm_Wm82_16446*	frameshift deletion	40912663	40912663	T	-	exon8	*Glyma.06g239000*
Chr06	*Gm_Wm82_16446*	nonframeshift insertion	40912665	40912665	-	CCC	exon8	*Glyma.06g239000*
Chr06	*Gm_Wm82_16446*	frameshift deletion	40912678	40912679	CC	-	exon8	*Glyma.06g239000*
Chr06	*Gm_Wm82_16446*	frameshift insertion	40912680	40912680	-	TA	exon8	*Glyma.06g239000*
Chr06	*Gm_Wm82_16446*	frameshift deletion	40912702	40912703	CC	-	exon8	*Glyma.06g239000*
Chr06	*Gm_Wm82_16712*	nonframeshift insertion	45704799	45704799	-	TTCCTTTATTTC	exon5	*Glyma.06g256000*
Chr06	*Gm_Wm82_16712*	frameshift insertion	45705118	45705118	-	TT	exon5	*Glyma.06g256000*
Chr06	*Gm_Wm82_16712*	frameshift insertion	45705122	45705122	-	C	exon5	*Glyma.06g256000*
Chr06	*Gm_Wm82_16712*	frameshift insertion	45705123	45705123	-	GGAC	exon5	*Glyma.06g256000*
Chr06	*Gm_Wm82_16712*	nonframeshift deletion	45705141	45705152	GAGAGTGTTTCA	-	exon5	*Glyma.06g256000*
Chr06	*Gm_Wm82_16712*	frameshift deletion	45708235	45708235	T	-	exon2	*Glyma.06g256000*
Chr06	*Gm_Wm82_16712*	nonframeshift deletion	45708244	45708246	TAT	-	exon2	*Glyma.06g256000*
Chr06	*Gm_Wm82_16714*	nonframeshift deletion	45717194	45717202	TGGAATCAG	-	exon4	*Glyma.06g256200*
Chr06	*Gm_Wm82_16771*	frameshift insertion	46527972	46527972	-	A	exon1	*Glyma.06g258900*
Chr06	*Gm_Wm82_16771*	frameshift insertion	46527996	46527996	-	CCGG	exon1	*Glyma.06g258900*
Chr06	*Gm_Wm82_16771*	nonframeshift deletion	46528040	46528066	ACCAACGACGGAAACCTCGTCCTTCTC	-	exon1	*Glyma.06g258900*
Chr06	*Gm_Wm82_16773*	nonframeshift deletion	46559916	46559918	TAG	-	exon4	-
Chr06	*Gm_Wm82_16773*	frameshift deletion	46559921	46559925	GAACA	-	exon4	-
Chr06	*Gm_Wm82_16773*	frameshift insertion	46559928	46559928	-	TT	exon4	-
Chr06	*Gm_Wm82_16773*	stopgain	46560013	46560013	-	ACTATTTCAAAATACGATTCATAATC	exon4	-

Chr06	*Gm_Wm82_16773*	frameshift deletion	46560017	46560020	CGCC	-	exon4	-
Chr06	*Gm_Wm82_16773*	frameshift insertion	46560024	46560024	-	TC	exon4	-
Chr06	*Gm_Wm82_16773*	nonframeshift deletion	46675775	46675777	GAA	-	exon1	-
Chr06	*Gm_Wm82_16810*	stopgain	47272370	47272370	-	GAAGGGCTATAGGGACTCACAGAGTACCACGA	exon1	*Glyma.06g263800*
Chr06	*Gm_Wm82_16810*	frameshift deletion	47275225	47275225	C	-	exon9	*Glyma.06g263800*
Chr06	*Gm_Wm82_16815*	nonframeshift deletion	47401043	47401045	GAA	-	exon2	*Glyma.06g265200*
Chr06	*Gm_Wm82_16819*	nonframeshift deletion	47437843	47437845	AAT	-	exon1	*Glyma.06g265400*
Chr06	*Gm_Wm82_16819*	frameshift insertion	47437861	47437861	-	GT	exon1	*Glyma.06g265400*
Chr06	*Gm_Wm82_16819*	frameshift deletion	47437865	47437866	GA	-	exon1	*Glyma.06g265400*
Chr06	*Gm_Wm82_16864*	nonframeshift deletion	48079792	48079806	TGGGAAAGCCCCATA	-	exon1	-
Chr06	*Gm_Wm82_16867*	nonframeshift insertion	48118225	48118225	-	GATGGCGGAGTCGGCATA	exon1	-
Chr06	*Gm_Wm82_16867*	frameshift deletion	48118266	48118270	CCGAA	-	exon1	-
Chr06	*Gm_Wm82_16867*	frameshift deletion	48118273	48118274	CG	-	exon1	-
Chr06	*Gm_Wm82_16867*	frameshift deletion	48118276	48118283	CATGAATG	-	wholegene	-
Chr06	*Gm_Wm82_16868*	nonframeshift deletion	48129505	48129516	CTGTTCAACCAT	-	exon3	*Glyma.06g270600*
Chr06	*Gm_Wm82_16871*	frameshift deletion	48154220	48154220	C	-	exon1	*Glyma.06g270900*
Chr06	*Gm_Wm82_16881*	frameshift insertion	48313892	48313892	-	T	exon5	-
Chr06	*Gm_Wm82_16888*	frameshift deletion	48447214	48447214	T	-	exon1	-
Chr06	*Gm_Wm82_16922*	nonframeshift deletion	48845564	48845572	CGGGTGGTG	-	exon1	*Glyma.06g276000*
Chr06	*Gm_Wm82_16923*	frameshift deletion	48846993	48847123	GGGTTTTGCATGGTGTAGTGAACGTCGTCCTGGGTGGCAAAGTGTATGACATGAGTGAAGAG	-	exon2	-

Chr08	*Gm_Wm82_23219*	nonframeshift deletion	43338186	43338203	GATGACGTCATTGAGG TTAACATCGACAT CGAAGATGTCGGTGC AATCGAGGAGGTG CTGGCAGGGGCG AAGCCTATTCCCATAGT T	-	exon4	*Glyma.08g293800*
Chr08	*Gm_Wm82_23231*	nonframeshift deletion	43486154	43486156	TTT	-	exon1	*Glyma.08g295400*
Chr08	*Gm_Wm82_23273*	frameshift insertion	44159176	44159176	-	TT	exon1	*Glyma.08g299000*
Chr08	*Gm_Wm82_23277*	frameshift deletion	44220686	44220687	AT	-	exon4	*Glyma.08g299600*
Chr08	*Gm_Wm82_23298*	frameshift deletion	44410462	44410462	G	-	exon6	*Glyma.08g301700*
Chr08	*Gm_Wm82_23305*	frameshift deletion	44533451	44533463	GTTCGACTGCTTC	-	exon5	*Glyma.08g302800*
Chr08	*Gm_Wm82_23307*	frameshift deletion	44561351	44561351	T	-	exon2	*Glyma.08g303000*
Chr08	*Gm_Wm82_23307*	frameshift deletion	44561530	44561563	TTATTCAGGCAAGGAAG ACATCCTTTGGCTTTGG	-	exon3	*Glyma.08g303000*
Chr08	*Gm_Wm82_23307*	frameshift deletion	44561635	44561647	ATGATCTGTCTTT	-	exon3	*Glyma.08g303000*
Chr08	*Gm_Wm82_23310*	frameshift deletion	44583701	44583702	CT	-	exon5	-
Chr08	*Gm_Wm82_23312*	nonframeshift insertion	44596176	44596176	-	GACGAT	exon1	*Glyma.08g303500*
Chr08	*Gm_Wm82_23312*	nonframeshift deletion	44596188	44596190	CAA	-	exon1	*Glyma.08g303500*
Chr08	*Gm_Wm82_23312*	frameshift deletion	44596198	44596201	TCCA	-	exon1	*Glyma.08g303500*
Chr08	*Gm_Wm82_23312*	frameshift insertion	44596203	44596203	-	G	exon1	*Glyma.08g303500*
Chr08	*Gm_Wm82_23312*	frameshift deletion	44596210	44596210	C	-	exon1	*Glyma.08g303500*
Chr08	*Gm_Wm82_23312*	frameshift deletion	44596212	44596213	CC	-	exon1	*Glyma.08g303500*
Chr08	*Gm_Wm82_23312*	frameshift insertion	44596222	44596222	-	CT	exon1	*Glyma.08g303500*
Chr08	*Gm_Wm82_23312*	frameshift insertion	44596224	44596224	-	TCCC	exon1	*Glyma.08g303500*
Chr08	*Gm_Wm82_23320*	frameshift deletion	44718339	44718339	C	-	exon3	*Glyma.08g304300*

Chr08	*Gm_Wm82_23352*	frameshift deletion	44995822	44995896	ATGGGTTCAATCCTCTC TACTAACAAAATCT AACATTTGCAAATTAA AAAAAAAAAAAA AAAAACTCAACCTGTT	-	exon1	*Glyma.08g307400*
Chr08	*Gm_Wm82_23371*	nonframeshift insertion	45253302	45253302	-	GGT	exon2	-

[1]Data not available

Supplementary Table 6. A list of QTL previously identified for sucrose content in soybean.

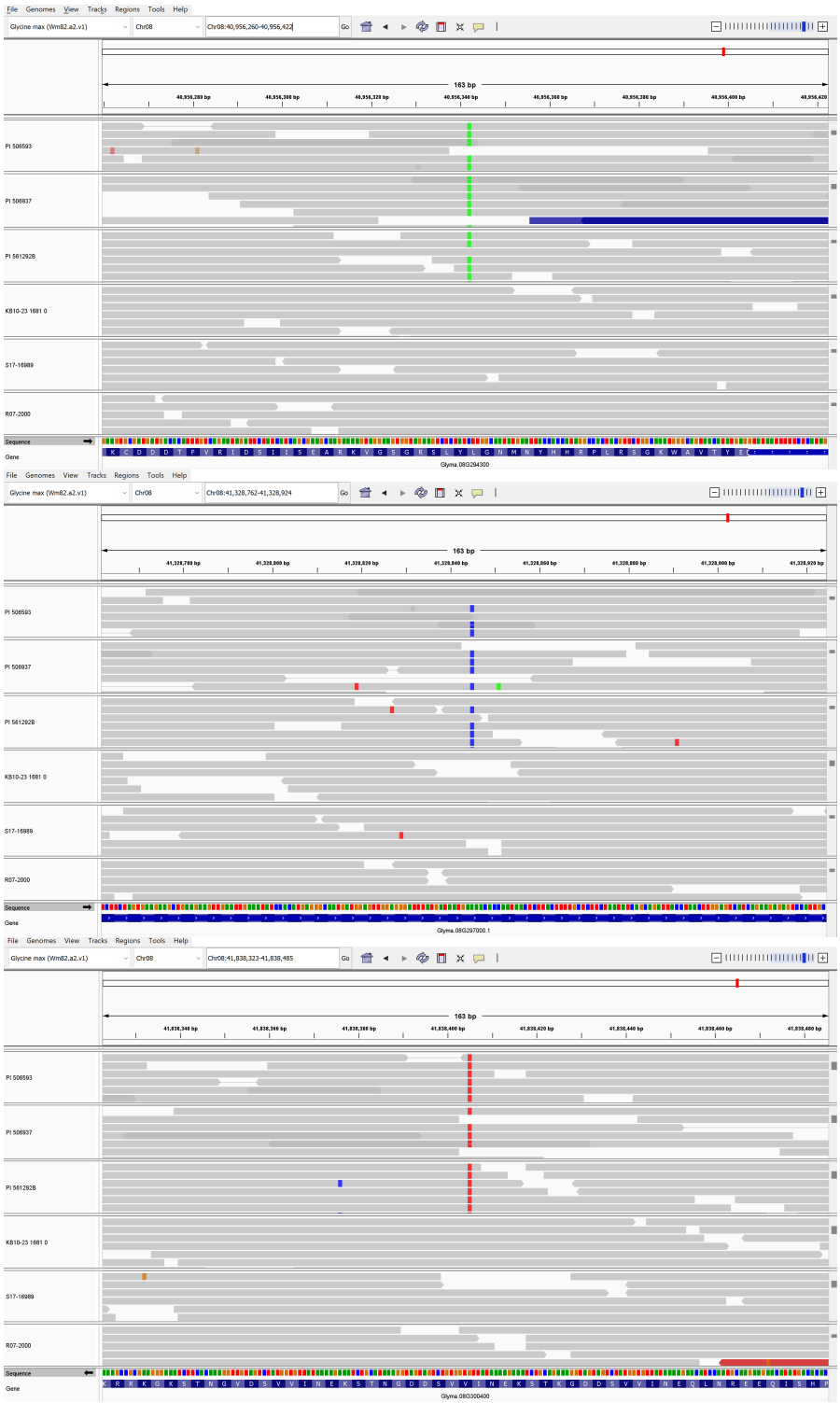
Chr	Soybase ID	Linked / Flanking Markers	Marker Position (bp)	PVE	Pedigree	Reference
Chr01	-[1]	Gm01_43979136	43,979,136	10.0	IA3023 × LD02-4485	Salari et al., 2021
Chr01	-	Gm01_3466825 / Gm01_3504836	3,466,825 - 3,504,836	20.5	Forrest × Williams 82	Knizia et al., 2023
Chr02	Sucrose 2-2	Satt546	39,547,350 - 41,404,328	-	Keunolkong × Iksan 10	Kim et al., 2005
Chr02	Sucrose 3-2	Satt537	17,770,710 - 31,234,716	3.9	Keunolkong × Shinpaldalkong	Kim et al., 2006
Chr02	-	Gm02_5155733 / Gm02_9925870	5,155,733 - 9,925,870	47.9	Forrest × Williams 82	Knizia et al., 2023
Chr02	-	Gm02_1199805 / Gm02_1373746	1,199,805 - 1,373,746	3.6	Forrest × Williams 82	Knizia et al., 2023
Chr03	-	ss244980861 / ss245114421	-	8.0	MD96-5722 × Spencer	Akond et al., 2015
Chr03	-	Gm03_192792	192,792	22.0	IA3023 × LD02-4485	Salari et al., 2021
Chr03	-	Gm03_1061417	1,061,417	22.0	IA3023 × LD02-4485	Salari et al., 2021
Chr03	-	Gm03_4113546 / Gm03_4595422	4,113,546 - 4,595,422	20.5	Forrest × Williams 82	Knizia et al., 2023
Chr04	-	Gm04_7672403	7,672,403	37.5	Forrest × Williams 82	Knizia et al., 2023
Chr05	Sucrose 1-1	A487_1	3,924,139 - 4,279,362	8.1	V71-370 × PI 407162	Maughan et al., 2000
Chr05	Sucrose 1-11	T169_1	-	7.0	V71-370 × PI 407162	Maughan et al., 2000
Chr05	-	Gm05_3273418 / Gm05_3867435	3,273,418 - 3,867,435	17.5	Forrest × Williams 82	Knizia et al., 2023
Chr05	-	Gm05_3748078 / Gm05_3803682	3,748,078 - 3,803,682	0.03	Forrest × Williams 82	Knizia et al., 2023
Chr05	Sucrose 4-1	ss245668753	14,845,260 - 20,147,198	46.0	MFS-553 × PI 243545	Zeng et al., 2014

Chr06	-	Sat213 / Satt643	14,644,152 - 16,106,552	76.0	PI 87013 × PI 200508	Skoneczka et al., 2009
Chr06	-	Sat213 / Satt643	14,644,152 - 16,106,552	60.0	PI 243545 × PI 200508	Skoneczka et al., 2009
Chr06	-	NCSB_001275 / NCSB_001285	36,800,000 - 38,000,000	9.2	Williams 82 × PI 483460B	Patil et al., 2018
Chr06	-	BARC_061709_17355 / SNP30079	36,700,000 - 38,000,000	12.4	Williams 82 × PI 483460B	Patil et al., 2018
Chr06	-	SNP35899 / BARC_014203_02694	18,600,000 - 19,100,000	8.5	Williams 82 × PI 483460B	Patil et al., 2018
Chr06	-	Gm06_1737718 / Gm06_5014399	1,737,718 - 5,014,399	10.5	Forrest × Williams 82	Knizia et al., 2023
Chr06	-	-	43,709,851 - 44,392,137	9.1	Hefeng 25 × Zhongdou 27	Liu et al., 2023
Chr07	Sucrose 1-9	GMSC514	167,990 - 509,374	7.3	V71-370 × PI 407162	Maughan et al., 2000
Chr07	-	Satt680	21,332,847 - 21,333,188	4.6	V97-3000 × V99-5089	Wang et al., 2014
Chr07	-	AX-90425348 / AX-90424508	142,962 - 222,695	17.2	Keunolkong × Iksan 10	Lee et al., 2016
Chr07	-	-	414,946 - 581,698	7.6	Hefeng 25 × Zhongdou 27	Liu et al., 2023
Chr08	Sucrose 1-2	A136_1	10,865,328 - 13,126,779	8.7	V71-370 × PI 407162	Maughan et al., 2000
Chr08	Sucrose 1-3	T153_1	7,892,162 - 8,937,354	7.6	V71-370 × PI 407162	Maughan et al., 2000
Chr08	Sucrose 1-12	A458_3	-	7.7	V71-370 × PI 407162	Maughan et al., 2000
Chr08	Sucrose 1-13	A486_1	8,283,676 - 9,192,408	8.4	V71-370 × PI 407162	Maughan et al., 2000
Chr08	-	AX-90497857 / AX-90396735	43,602,528 - 43,670,249	15.9	Keunolkong × Iksan 10	Lee et al., 2016
Chr08	-	NCSB_001722 / BARC_035233_07151	9,040,000 - 9,390,000	22.6	Williams 82 × PI 483460B	Patil et al., 2018
Chr08	-	SNP16086 / BARC_016683_03318	8,510,000 - 8,720,000	35.4	Williams 82 × PI 483460B	Patil et al., 2018
Chr08	-	Gm08_5960619 / Gm08_8268861	5,960,619 - 8,268,861	0.04	Forrest × Williams 82	Knizia et al., 2023
Chr09	-	ss246864598 / ss246876159	-	8.0	MD96-5722 × Spencer	Akond et al., 2015
Chr09	-	Gm09_1888876	1,888,876	20.5	Forrest × Williams 82	Knizia et al., 2023
Chr09	-	-	22,562,116 - 23,609,852	6.7	Hefeng 25 × Zhongdou 27	Liu et al., 2023
Chr09	Sucrose 4-2	ss246796276	2,967,367 - 5,901,485	10.0	MFS-553 × PI 243545	Zeng et al., 2014
Chr10	-	Gm10_621706	621,706	19.1	Forrest × Williams 82	Knizia et al., 2023

Chr11	Sucrose 2-1	Satt197	8,602,082 - 9,106,824	-	Keunolkong × Iksan 10	Kim et al., 2005
Chr11	-	Satt359	32,411,307 - 32,411,478	13.4	V97-3000 × V99-5089	Wang et al., 2014
Chr12	Sucrose 3-4	Satt442	6,653,096 - 7,533,328	8.3	Keunolkong × Shinpaldalkong	Kim et al., 2006
Chr13	-	-	38,932,430 - 39,299,827	6.3	Hefeng 25 × Zhongdou 27	Liu et al., 2023
Chr13	Sucrose 1-5	A186_1	26,196,486 - 28,912,864	9.6	V71-370 × PI 407162	Maughan et al., 2000
Chr13	Sucrose 1-16	OO20850	-	6.1	V71-370 × PI 407162	Maughan et al., 2000
Chr13	-	Gm13_3524828 / Gm13_3891723	3,524,828 - 3,891,723	17.5	Forrest × Williams 82	Knizia et al., 2023
Chr14	Sucrose 3-1	Satt020	16,752,071 - 38,859,467	7.1	Keunolkong × Shinpaldalkong	Kim et al., 2006
Chr15	Sucrose 1-10	A963_1	5,104,450 - 5,215,740	6.9	V71-370 × PI 407162	Maughan et al., 2000
Chr15	Sucrose 3-3	Satt483	13,755,345 - 17,021,739	14.8	Keunolkong × Shinpaldalkong	Kim et al., 2006
Chr15	-	ss248604753 / ss248616287	-	68.0	MD96-5722 × Spencer	Akond et al., 2015
Chr16	Sucrose 3-5	Satt287	1,357,034 - 2,827,903	4.1	Keunolkong × Shinpaldalkong	Kim et al., 2006
Chr16	Sucrose 4-3	ss249186914	-	8.0	MFS-553 × PI 243545	Zeng et al., 2014
Chr16	Sucrose 3-6	Sct_065	4,952,453 - 6,325,509	8.3	Keunolkong × Shinpaldalkong	Kim et al., 2006
Chr17	-	Gm17_4967175 / Gm17_5294475	4,967,175 - 5,294,475	20.5	Forrest × Williams 82	Knizia et al., 2023
Chr17	-	-	4,517,109 - 7,005,804	16.8	Hefeng 25 × Zhongdou 27	Liu et al., 2023
Chr18	-	Gm18_1620585 / Gm18_2020823	1,620,585 - 2,020,823	17.5	Forrest × Williams 82	Knizia et al., 2023
Chr19	Sucrose 1-6	A023_1	35,672,961 - 36,154,603	9.7	V71-370 × PI 407162	Maughan et al., 2000
Chr19	Sucrose 1-7	B164_1	33,865,153 - 36,578,530	9.6	V71-370 × PI 407162	Maughan et al., 2000
Chr19	Sucrose 1-8	B162_2	40,205,349 - 40,265,091	7.0	V71-370 × PI 407162	Maughan et al., 2000
Chr19	Sucrose 1-17	OB41600	-	8.6	V71-370 × PI 407162	Maughan et al., 2000
Chr19	Sucrose 2-3	Satt523	4,244,065 - 12,744,826	-	Keunolkong × Iksan 10	Kim et al., 2005
Chr19	Sucrose 2-4	Satt278	12,234,826 - 34,351,864	-	Keunolkong × Iksan 10	Kim et al., 2005
Chr19	Sucrose 2-5	Satt418	12,234,826 - 34,059,981	-	Keunolkong × Iksan 10	Kim et al., 2005

Chr19	Sucrose 2-6	Satt398	9,284,015 - 34,059,981	-	Keunolkong × Iksan 10	Kim et al., 2005
Chr19	Sucrose 2-7	Satt462	37,425,576 - 38,937,316	-	Keunolkong × Iksan 10	Kim et al., 2005
Chr19	Sucrose 2-8	Satt497	34,351,864 - 36,154,603	-	Keunolkong × Iksan 10	Kim et al., 2005
Chr19	Sucrose 2-9	Satt313	34,351,864 - 35,280,772	-	Keunolkong × Iksan 10	Kim et al., 2005
Chr19	Sucrose 2-10	Sct_010	40,637,071 - 41,616,190	-	Keunolkong × Iksan 10	Kim et al., 2005
Chr19	Sucrose 2-11	Satt076	40,637,071 - 41,616,190	-	Keunolkong × Iksan 10	Kim et al., 2005
19	-	AX-90508115 / AX-90509468	34,327,636 - 34,692,448	36.9	Keunolkong × Iksan 10	Lee et al., 2016
19	-	-	45,311,975 - 45,464,136	13.2	Hefeng 25 × Zhongdou 27	Liu et al., 2023
20	Sucrose 1-4	A144_1	2,716,974 - 25,498,552	12.4	V71-370 × PI 407162	Maughan et al., 2000
20	Sucrose 1-14	K227_4	-	11.4	V71-370 × PI 407162	Maughan et al., 2000
20	Sucrose 1-15	nbs37	-	11.3	V71-370 × PI 407162	Maughan et al., 2000
20	-	Satt270	35,362,576 - 35,362,794	5.2	V97-3000 × V99-5089	Wang et al., 2014
20	-	Gm19_2552468	2,552,468	9.1	Forrest × Williams 82	Knizia et al., 2023
20	-	-	701,169 - 807,404	5.7	Hefeng 25 × Zhongdou 27	Liu et al., 2023
20	-	-	40,523,599 - 41,882,459	14.4	Hefeng 25 × Zhongdou 27	Liu et al., 2023

[1]Data not available



Supplementary Figure 1. The unique SNP variants identified in this study in other high sucrose germplasms and mutant lines.

VITA

Dongho Lee was born on March 2, 1990 and raised in Gyeongju, South Korea. Dongho Lee obtained his bachelor's degree in Plant Biosciences at Kyungpook National University, Daegu, South Korea in 2015. During his bachelor's degree, he first came to the United States as an undergraduate research intern at the University of Missouri – Fisher Delta Research, Extension, and Education Center under Dr. James Grover Shannon's supervision in July 2013 – December 2013. In 2017, Dongho Lee obtained his master's degree in Plant Biosciences, emphasizing soybean breeding and genetics under Dr. Jeong-Dong Lee at Kyungpook National University. After graduating with a master's degree, Dongho Lee returned to the University of Missouri – Fisher Delta Research, Extension, and Education Center and worked as a post-graduate research assistant until August 2019. In 2019, Dongho Lee joined Dr. Pengyin Chen's soybean breeding program at the University of Missouri – Fisher Delta Research Extension, and Education Center to pursue a PhD, emphasizing soybean genetics and genomics.