UNDERSTANDING SUCROSE AND RAFFINOSE FAMILY OF OLIGOSACCHARIDES IN SOYBEAN SEED FOR HUMAN AND ANIMAL UTILIZATION

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

HYUN JO

Dr. Kristin Bilyeu, Dissertation Supervisor

June 2016

The undersigned, appointed by the dean of the Graduate School,

have examined the dissertation entitled

UNDERSTANDING SUCROSE AND RAFFINOSE FAMILY OF OLIGOSACCHARIDES IN SOYBEAN SEED FOR HUMAN AND ANIMAL UTILIZATION

Presented by Hyun Jo

A candidate for the degree of

Doctor of Philosophy

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

Dr. Kristin Bilyeu

Dr. James Schoelz

Dr. Andrew Scaboo

Dr. Fu-Hung Hsieh

DEDICATION

I would like to dedicate this dissertation to my parents. Without their inspiration, I would never think about doctoral course in my life. Thank you so much for all of your guidance, supports, and confidence in me. Also, I appreciate all the helps from my wife, Eunhae Park. Because you were with me, I was able to complete each step to achieve the degree. Finally, thank God that he always has helped me to move forward and that he has allowed such wonderful people around me.

ACKNOWLEDGEMENTS

I cannot imagine my Ph.D life without my great advisor Dr. Kristin Bilyeu. She was not only my academic advisor but also my mentor of my life. Her warm but strong heart always encouraged me a lot. Without her help and guidance, I could not have completed this degree. Whenever I lost she shed a light to me. It has been my fortune to have Dr. Kristin Bilyeu as my advisor.

Also, this study would not have been possible without supports and guidance of my committee members, Dr. Fu-Hung Hsieh, Dr. James Schoelz, and Dr. Andrew Scaboo. Thank you so much for all of their helps. I was able to be on the right path to complete the study because they shared their wisdom with me. I am honored to have them as my committee.

I also want to express my sincere thanks to technicians of Dr. Kristin Bilyeu's lab, Paul Little and Christine Cole. Whenever I have questions about specific experiments I need to follow in the lab or field, their answers were the key to find a right direction. Especially, I would also like to thank Dr. Jeong-dong Lee and Dr. Jason Gilman who helped me to carry out data analysis. Thank you so much for helping me during this program.

Besides academic people, my friends were the ones I can depend on during this tough journey as a doctoral student. I feel like I met so many good lifetime friends. I would like to express special thanks to Surya Sapkota, Carrie Miranda, and Dr. Yoonha Kim.

Last but not least, thank God that he allows me to have all the great people around me. My program period as a doctoral student were blessed by him.

TABLE OF CONTENTS

LIST OF TABLES

Table **Page**

LIST OF FIGURES

Figure **Page**

Chapter 1

Literature review

Cultivated soybean [*Glycine max* (L.) Merr.] is a member of the Fabaceae family and the *Glycine* genus, which includes about 20 annual species. The wild soybean (*Glycine soja* Sieb. and Zucc.) is considered an ancestor of modern soybean cultivars. Cultivated soybean and wild soybean are included in the subgenus *Soja* which is diploid (2n = 40).

Domestication and dissemination of soybean from the region of origin

Domestication of modern soybean occurred in northeast China around the 11th century B.C during the Shang dynasty or earlier. The conversion from wild species to cultivated species occurred through selections made by farmers (Hymowitz 1970; Hymowitz and Newell 1980). Although the dissemination of soybean is not fully known, scholars suggest that soybean was widely grown in China and the Korean peninsula by the first century A.D. Soybean was disseminated primarily through trade routes from China throughout Asian countries such as Japan, Indonesia, Vietnam, Thailand, Nepal, India, Philippines, Myanmar and Malaysia by the 15th century. Since China and Korea are the center of origin for domesticated soybean, those countries are considered to be secondary gene centers of soybean (Hymowitz 2008). Soybean has become the source of important food materials for Asian counties such as miso, soy sauce and tofu.

By the 18th century, soybean was introduced into Europe. Linnaeus mentions in the *Hortus Cliffortianus* that soybean was introduced into The Netherlands before 1737. Hymowitz and Harlan (1983) stated that soybean was planted in 1740 in France and in 1790 in England. Samuel Bowen, a seaman in the East India Company, is the first person to bring soybean to the Colony of Georgia from China via Europe in 1765 (Hymowitz and Harlan 1983). Bowen had

grown soybeans on his planation and used them to make soy sauce and vermicelli (soybean noodles) (Hymowitz, 1990; Hymowitz and Shurtleff 2005). Benjamin Franklin sent soybean seeds from Europe to Philadelphia in 1770, another introduction of soybean to U.S. However, after the introduction of soybean to the U.S. by Bowen, it took almost 150 years for soybean to be widely grown in the U.S. (Singh and Hymowitz, 1999). For many years, soybean was mainly used in the U.S. as forage for livestock instead of harvesting the soybean seed. Eventually, researchers were able to show that cooked soybean, as opposed to raw soybeans, had beneficial nutritional properties for humans. Thus, as soybean consumption has increased, many researchers have become interested in the development of soybean lines for human consumption, as well as for the processing industry (Hymowitz, 1990). Finally, the U.S. has been become the world's highest producer. In addition, as they have known the importance of having and maintaining a set of genetically diverse plants, and accessions from other countries were collected for the U.S. To date, the U.S. germplasm collection includes more than 21,000 *Glycine* accessions.

Economic importance of soybean

Soybean is an important annual crop, providing edible oil (about 20 %), protein (about 40 %) and soluble carbohydrate (about 20 %) for both human and livestock consumption. In 2012, the highest U.S. vegetable oil consumption was from soybean, higher than consumption of oils produced from canola, palm or coconut. About 85 % of the soybean grown in the world is produced in North (U.S. and Canada) and South America (primarily Argentina and Brazil). The area devoted to soybean production in South America has increased to more than 35 million

hectares, whereas areas in the U.S. and Canada have remained at about 30 million hectares over the last 15 years (www.soystats.com for 2015). The demand for soybean products has also risen with increases in the world's population and in the need for animal feeds. This in turn has meant that the value of the soybean crop had continually risen since 2005 and has decreased last three years. In addition, soybeans are also used for other purposes such as industrials (e.g. varnish and soap) and pharmaceutical products. In Asian countries, traditional soybean-foods such as natto, soybean milk, soy sauce, soybean- paste and soybean sprouts are a very popular food that are consumed daily. In the U.S., the two main products are oil and protein-containing meal.

In 2014, soybean production in the U.S. was 3,969 million bushels. Total value of the U.S. soybean crop was nearly $40.3 billion. The U.S. produced 18.3 billion pounds of soybean oil which represented 59 % of the world's oilseed production and 43.1 million short tons of soybean meal. Approximately 55 % of the vegetable oil consumed in the U.S. was from soybean. Edible oil consumption of soybean was used for salad and cooking oil, shortening and margarine. Soybean meal is mainly used as feed for animals. Poultry consumes the majority of soymeal, using 55 % of U.S. soymeal, followed by swine and beef, which consume 22 % and 11 %, respectively (SoyStats 2015).

Seed composition in soybean

Soybean seeds typically contain 40 % protein, 20 % oil, 35 % carbohydrates and 5 % ash on a dry weight basis (Liu, 1997). Generally, soybean protein is approximately 40 %, but almost 50 % protein content is found in soybean lines present in the world soybean collections. Of the total protein, 70-80 % is comprised of the storage proteins known as glycinin and ß-conglycinin

(Verma and Shoemaker, 1996). Although those proteins have no catalytic function, they are important components that determine the nutritional properties of foods from soybeans. The other storage proteins have catalytic activity in soybean including lipoxygenases, trypsin inhibitors, lectins and urease (Verma and Shoemaker, 1996). These storage proteins can function to deter potential insect and animal predators, and those components are anti-nutritional factors for making food and feeds. Although soybean seed has high protein, protein in soybean seed is negatively correlated with soybean yield (Verma and Shoemaker, 1996).

Soybean carbohydrates are classified as structural and non-structural, and they accumulate to approximately 35 % of soybean seed (Verma and Shoemaker, 1996). Structural carbohydrates include fiber components such as cellulose, hemicelluloses, pectins and glycoproteins (Bach Knudsen et al., 1987; Middelbos and Fahey Jr. 2008; Selvendran et al., 1987). In non-structural carbohydrates, there are starch, sucrose and oligosaccharides (raffinose and stachyose) that total about 12 % of the dry seed weight (Karr-Lilienthal et al., 2005; Wilson, 2004). Sucrose from photosynthate in leaves is translocated into seed. Accumulation of sucrose is ceased when the color of seeds is changed from green to yellow during seed maturation, whereas raffinose and stachyose increase after cessation of sucrose accumulation (Obendorf et al., 1997). Mostly, the mature soybean seed has 41 – 68 % sucrose, 5 – 16 % raffinose and 12 – 35 % stachyose of soluble carbohydrates (Verma and Shoemaker, 1996). The highest level of starch detected during the seed developmental process is 11-12 %. Starch content dropped to less than 3 % by the end of this period (Yazdi-Samadi et al., 1997; Wilson, 1987).

Soybean oil accounts for the major part of oil production in the world. This component of soybean seed is found in the cotyledons of seed (Gerde and White, 2008). Oil consists of

triacylglycerol with three of the same or different fatty acids bonded to the glycerol backbone. The fatty acid composition determines oil quality. Soybean oil has five common fatty acids such as 11 % palmitic (16:0), 4 % stearic (18:0), 23 % oleic (18:1), 54 % linoleic (18:2) and 8 % linolenic (18:3) (Fehr, 2007; Wilson 2004). Palmitic and stearic acid are saturated fatty acid as they do not contain any double bonds and the carbons are instead saturated with hydrogen bonds. On the other hand, oleic, linoleic and linolenic are unsaturated fatty acids which contain double bonds. One of the key goals to improve soybean oil quality is to reduce linolenic (18:3) fatty acid to increase oxidative stability. According to the position of double bonds, there are ω-6 and ω-3 types of fatty acids, which are essential to humans and animals.

Soy foods

Soybean for food is mainly focused on consumption of high protein food and vegetable oil as its characteristics. In comparison with the expense of producing beef, soybean is an inexpensive alternative to produce high protein food. Traditionally, soybean has been used as a staple food in Asia as soymilk, tofu and soy sauce (Liu, 1997; Watanabe, 1997). In addition, increased recognition of the health benefits of soy foods has increased soy food consumption in North America instead of consumption of animal products (Murphy et al., 1997).

Soymilk is made by processing soybean that has been soaked overnight, ground with added water, strained and heating. Soymilk is a good substitute for plain milk and an excellent source of high quality protein and B-vitamins (www.soyfoods.com). In addition, there is no trypsin inhibitor after the heating process. And as soybean has no galactose and lactose, soymilk can be a good alternative to breast milk and lactose-free milk. Tofu, which is like a soft cheese

food, is produced from soy milk by curdling and then draining (www.soyfoods.com). As tofu has a high amount of protein and low calories, it often used as a meat substitute in Asian diets in the past. Thus, a lot of vegetarians take tofu as a source of non-animal protein food in Western countries. Soy pulp remains after the process of making soymilk and tofu. It is called okara which is high-quality, inexpensive sources of dietary fiber. Of course, okara has less protein than seed and also can be used as a food ingredient for cuisines in Korea and Japan (www.soyfoods.com).

Miso and Natto are traditional Japanese foods. Miso is a rich, salty condiment made from soybeans and a grain such as rice with salt and a mold culture. Miso is used as flavor for soups, sauces, dressing and marinades (www.soyfoods.com). Natto is made of fermented, cooked whole soybeans. Because the fermentation process breaks down the beans' complex proteins, natto is more easily digested than whole soybeans (www.soyfoods.com). Soy sauce is a dark brown liquid made from soybeans by a fermenting process. Soy sauces have a salty taste, but are lower in sodium than table salt. There are specific types of soy sauces such as shoyu, tamari and teriyaki (www.soyfoods.com).

Whole soybeans can be soaked in water and then baked until browned, and this soy food is referred to as roasted soynuts. Soynuts are rich in protein and isoflavone and similar in texture and flavor to peanuts (www.soyfoods.com). Edamame is a green vegetable soybean food as a snack or a main vegetable dish. The large soybeans are harvested when they are still green. They retain their sweet taste after boiling in slightly salted water. They are high in protein and fiber and contain no cholesterol (www.soyfoods.com).

Soybean sprouts are grown under dark conditions about 1-2 weeks with frequent watering and produce a soy food that is an excellent source of nutrition, packed with protein and vitamin C. They can also be used raw in salads or soups, or in stir-fried, sautéed, or in baked dishes (www.soyfoods.com). Soybean oil is the natural oil from whole soybeans. It is the most widely used oil in the U.S., accounting for more than 75 percent of our total vegetable fats and oils intake. Soybean oil is cholesterol-free and high in polyunsaturated fat. Soybean oil also is used to make margarine and shortening (www.soyfoods.com). Soy flour is made from roasted soybeans ground into a fine powder. Other soyfoods include soy yogurt, yuba, soynut butter and soy nuggets.

Soy feeds

Approximately 80 % of the soybean meal production is used in animal feeds in the U.S. Soybean-based animal feeds are mostly consumed in poultry, swine, beef, dairy and fish production systems. Although soybeans produce a high-quality cost-effective meal, it is necessary to supplement feed with essential amino acids, such as methionine, cysteine, and lysine to have optimum animal performance (Kwanyuen et al., 1997). Soybean meal is a beneficial source of protein and nutrients in animal feed.

Anti-nutritional factors in soybean seed

Some of the soybean seed constituents cannot be digested by animals or humans. Among these, trypsin inhibitors are the most important for soybean. Unprocessed soybeans contain high

levels of trypsin inhibitors, around 6 % of the total protein of the seed (Orf and Hymowitz, 1979). Heat treatment is an essential process to remove trypsin inhibitors for livestock feed, although this process is expensive (DiPietro and Liener, 1989). Studies have shown that heat treatment of soybean meal to lower the level of trypsin inhibitors increases the growth of chicks (Anderson-Hafermann et al., 1992) and channel cat fish (Wilson and Poe, 1985). However, the heating process can destroy the activity of other beneficial nutrients.

The availability of phosphorus is another concern for soybean seed quality. Approximately 75 % of phosphorus is stored as phytic acid in soybean seed (Raboy et al., 1984). Phytic acid is indigestible by non-ruminant animals and may contribute to chelation of essential minerals (Heaney et al., 1991; Zhou et al., 1992; Lynch et al., 1994) and phosphorus pollution in manure (Sebastian et al., 2000). Monogastric animals lack the enzyme phytase to mobilize this form of phosphorus (Raboy, 2007a). The negatively charged phytate has a high affinity for micronutrients such as Fe^{2+} and Zn^{2+} (Raboy, 2002). When unused phosphorus is released into fresh water or lakes, it causes environmental eutrophication (Correll, 1998; Raboy, 2007b).

A third anti-nutritional factor in soybean seed is in the form of soluble carbohydrates. Soybean seeds accumulate about 20 % of soluble carbohydrates. Among the different carbohydrates in the seed, the raffinose family of oligosaccharides (RFO) raffinose, stachyose, and verbascose are anti-nutritional carbohydrates. They are derived from sucrose by the subsequent addition of galactose. Chain elongation is catalyzed by raffinose synthase and stachyose synthase (Peterbauer and Richter, 2001). RFO cannot be digested by humans and monogastric animals due to the lack of α-galactosidase in the gut. However, RFO are microbially fermented in the lower gut resulting in flatulence and reduced feed efficiency (Gitzelmann and

Auricchio, 1965; Ruttloff et al., 1967; Price et al., 1988; Sebastian et al., 2000). Consumption of low-raffinose, low-stachyose soybean seed products reduced flatulence in humans (Suarez et al., 1999) and increased metabolizable energy efficiency in chickens (Parsons et al., 2000).

Soybean maturity group and architecture of soybean

Because soybean is a short-day crop, the development of soybean is largely determined by day-length, unlike many other crops that depend on temperature. This photoperiod sensitivity means that soybean will flower and reach reproductive development stages from vegetative development stages when the night time lengthens. Thus, soybean flowering is actually determined by the specific length of night time. It is important for farmers to consider maturity to determine seed production and quality in a specific latitude in U.S. production environments. Soybeans that grow for a full season have a high quality and yield compared to too early matured soybeans or too late matured seed. Crookston and Hill (1978) reported that seed shrinkage and loss of green pod color identified physiological maturity of individual soybean plants. The best time for soybean physiological maturity and harvest occurs right before the first significant frost.

There are several major soybean maturity categories in North America from Maturity Group (MG) 00, adapted to the border between U.S and Canada to MG VIII, adapted to the southern U.S. Early maturity groups require a short dark period to transition to reproductive development, whereas late maturity groups need a long period of darkness to initiate flowering. When MG 00 varieties have been grown in the southern U.S., there is not enough time for vegetative development and production is sharply reduced. By contrast, late maturity groups grown in the northern U.S. may be affected by frost damage. Varieties in a given maturity group

will usually perform as optimally full-season varieties within a band of similar latitude that is no wider than 100 to 150 miles from north to south.

As soybeans in the various maturity groups are related to plant growth type, there are mainly two stem architecture types: indeterminate and determinate. Soybeans grown in the North are mostly of the indeterminate growth type. With indeterminate varieties, vegetative and reproductive development occurs simultaneously for several weeks, which means that soybean continues to grow after flowering. Generally, varieties in Maturity Groups 00 through Group IV are indeterminate (Wilcox, 1987). On the other hand, most southern varieties are determinate. With determinate varieties, flowering occurs when vegetative development has ceased. In general, varieties in maturity Groups V through VIII are determinate (Wilcox, 1987).

Maturity genes in soybean

The development of cultivars with early maturity for the northern U.S. requires effective use of early maturity genes. Seven loci have been identified that influence the time to maturity. Alleles at these loci act on photoperiod sensitivity. Using these loci has historically been difficult, and improved molecular markers could assist the soybean breeder in selecting for early maturity (Molnar et al., 2003).

The maturity genes have been identified by classical methods. Several maturity loci designated as *E* loci have been found in soybean, including *E1* and *E2* (Bernard, 1971), *E3* (Buzzell, 1971), *E4* (Buzzel and Voldeng, 1980), *E5* (McBlain and Bernard, 1987), *E6* (Bonato and Vello, 1999), *E7* (Cober and Voldeng, 2001) and *E8* (Cober et al., 2010). Maturity gene *e6* is

dominant for late flowering while for the remaining *E* genes, the recessive version conditions early flowering. *E1* is located on linkage group (LG) C2 at position 113.0 cM on the 2003 composite map (Glyma06g23026) (Song et al., 2004). *E7* was linked to *E1* on LG C2 and was shown to be an independent allele of *E1* (Molnar et al., 2003; Xia et al., 2012). E1 has been cloned and is a novel gene that appears to encode a transcription factor (Xia et al., 2012). There are several allelic variations in the *E1* gene including a single nucleotide polymorphism (SNP) at 44 base pair position with a missense mutation (arginine to threonine) (*e1-as*), a single base deletion at 49 base pair position with a premature stop codon at 124 base pair position (*e1-fs*), and the entire *E1* gene deletion, null allele (e1-nl) (Xia et al., 2012). *E2* is located at 136.3 cM on LG O (Glyma10g36600) which is an orthologue of Arabidopsis GIGANTEA (GI) gene (Watanabe et al., 2011). A nonsense mutation was found at 1561 base pair when thymine base is substituted for adenine base (*e2-ns*) (Watanabe et al., 2011). *E3* is located on LG L (Glyma19g41210) (Molnar et al., 2003) which is a homolog of the photoreceptor phytochrome A (Liu et al., 2008; Franklin and Quail 2010). There are two functional *E3* alleles such as *E3* and *E3* short (Langewisch et al., 2014). Both *E3* alleles have identical protein sequences. However, the *E3* allele, which is identical to Williams 82 sequence, has an insertion in the third intron in comparison with the *E3* short allele. There is no evidence to prove any functional differences between those *E*3 alleles (Liu et al., 2008). Having a 13-kb deletion in the *E3* gene, the mutant *e3* allele does not contain the fourth exon (Liu et al., 2008). Another mutant *e3-ft3* allele has a missense mutation that glycine is substituted for arginine at 1050 position of the amino acid sequence (Liu et al., 2008). *E4* is located on LG I (Abe et al., 2003; Molnar et al., 2003) which is a homolog of *E3*. There is one functional *E4* allele and five variants reported to date. The mutant

e4 allele has a large insertion in exon 1 causing a premature stop codon that truncates the 1123 amino acid sequences to 237 amino acids (Liu et al., 2008). Four *e4* variants each have one single nucleotide deletion in the coding sequence making truncated amino acid sequences (Tsubokura et al., 2013).

The effect of Germination

Energy from stored carbohydrates in seeds is thought to be necessary for germination (Bewley and Black, 1994; Peterbauer et al., 2001). The RFO have an important role in seed protection as they are thought to stabilize membranes during seed desiccation (Bailly et al., 2001; Obendorf, 1997). Soybean seeds have a high concentration of stachyose, but the concentrations of raffinose and stachyose decrease 18 hours before germination (Hsu et al., 1973; Koster and Leopold, 1988). The loss of stachyose is associated with the loss of desiccation tolerance (Koster and Leopold 1998). The mutant myo-inositol phosphate synthase (*mips)* phenotype is associated with reduced raffinose, stachyose and phytin seed concentrations, and soybeans expressing the mutant *mips* phenotype have a low emergence in the field (Sebastian et al., 2000; Hitz et al., 2002). In addition, Blöchl et al. (2007) directly tested germination of pea seed by removing raffinose and stachyose metabolism, and the results indicated that germination was significantly delayed (Blöchl et al., 2007).

On the other hand, inhibition of RFO metabolism in soybean seeds did not decrease germination, suggesting that RFO metabolism is not a requirement for soybean germination (Dierking and Bilyeu, 2009). These results suggest that other components are necessary for germination. PI 200508 was the soybean plant introduction demonstrated to have a low level of

RFO controlled by the recessive allele *stc1a* (Neus et al., 2005). Since the *stc1a* allele was evaluated with agronomic traits such as emergence, seed yield, lodging and height; the authors concluded that there were no significant differences between *Stc1a* and *stc1a* lines (Neus et al., 2005). A soybean line, SGUL, with an ultra-low raffinose and stachyose phenotype, was studied for germination rate and the SGUL line had at least 83 % of germination rate compared to soybean lines not having a low raffinose and stachyose seed content (Schillinger et al., 2013).

Role of carbohydrates in transport and storage

Plants mainly use the disaccharide sucrose as well as the oligosaccharides raffinose, stachyose, and verbascose for transport. The use of sucrose and the RFO has several advantages for long-distance transport and storage carbohydrates for many plants. During translocation of RFO in plants, the phloem-mobile RFO are formed in specialized companion cells (Beebe and Turgeon 1992; Turgeon 1996).

Accumulation of RFO and activity of enzymes during seed development

During seed development, soybean seeds increase in dry weight and their color eventually changes from green to yellow as they mature. Soybean seeds begin to shrink from the pod as they lose water. Physiological maturity, which is an important time during seed maturation, is defined as the time of maximum seed dry weight. Seeds reach physiological maturity around 50 days after pollination and seed moisture is about 60 % of fresh weight basis (Obendorf et al., 1980). RFO usually accumulate in the late stages of seed development (Amuti and Pollard, 1977; Yazdi- Samadi et al., 1977; Lowell and Kuo, 1989). About 70 % of the RFO

accumulate after reaching physiological maturity during the phase of seed drying (Obendorf et al., 2009). The monogalactoside starts to accumulate when the embryo tissues begin to yellow, followed by the digalactosides and finally the trigalactosides during the desiccation step of seed maturation (Obendorf et al., 2009).

The exchange and synthesis reactions catalyzed by raffinose synthase were detected in soybean five days after flowering. However, while the exchange activity continued to increase up to 60 days flowering, the synthesis activity leveled off at 15 days after flowering (Castillo et al., 1990). In addition, enzyme activity for raffinose synthase was measured *in vitro* on harvested seed from the pod at the beginning to yellow stages, and enzyme measurement was decreased in a mutant line compared to the control line (Hitz et al., 2002).

The effect of environmental stresses

Environmental factors such as drought, heat and salinity are abiotic stresses that effect plant growth. Changing gene expression, plants grown under these stresses have low yield. Under these stresses, soluble carbohydrates, sugar alcohols, cyclic carbohydrates and RFO have an important role in stabilizing membranes and proteins (Tarczynski et al., 1993). Gene expression and enzyme activity involving RFO are associated with environmental stresses such as imbibitional chilling, cold, salinity and drought stress (Nishizawa et al., 2008).

Obendorf et al. (2008) indicated that soybean lines with low raffinose, stachyose, and phytin were sensitive to imbibitional chilling. In addition, the imbibitional chilling tolerance of the low raffinose and stachyose soybeans was speculated to be due to the higher accumulation of

cyclitol α-galactosides in the embryo (Obendorf et al., 2008). In leaves of tomato and Arabidopsis seedlings, raffinose synthase increased under environmental stresses such as cold and desiccation (Downie et al., 2003; Zuther et al., 2004). In addition, soluble sugar and RFO were accumulated during the process of cold acclimation (Gilmour et al., 2000; Koster and Lynch, 1992). With petunia, the accumulation of raffinose was related with increased freezing tolerance of the leaves (Pennycooke et al., 2003). In a drought study with Arabidopsis, the accumulation of raffinose is important for drought stress to maintain cell turgor and stabilize cell proteins (Taji et al., 2002). Raffinose can stabilize isolated chloroplast thylakoid membranes during a freeze–thaw cycle as well (Hincha, 1990). Plants with elevated raffinose under normal growth conditions have shown higher drought and freezing tolerance than wild type plants.

It is generally accepted that the imposition of environmental stresses such as drought, chilling, heat, and high-light irradiation give rise to excess concentrations of reactive oxygen species (ROS) and that much of the injury to plants caused by exposure to stress is associated with oxidative damage at the cellular level (Bowler et al., 1992; Alscher et al., 1997; Shigeoka et al., 2002). Based on Nishizawa et al., (2006), it seems that galactinol and RFO protect plant cells from the oxidative damage caused from various types of stresses.

Biosynthesis of Raffinose Family of Oligosaccharides

RFO are composed of raffinose, stachyose and verbascose. RFO are synthesized from sucrose by the subsequent addition of galactose from galactinol (Pharr et al., 1984). Galactinol synthase is the enzyme required for the first step of the RFO biosynthesis pathway, and it catalyzes the synthesis of galactinol from UDP-galactose and myo-inositol. Galactinol is used as

a galactosyl donor for raffinose, stachyose and verbascose (Senser et al., 1967). The trisaccharides (raffinose) is elongated from sucrose by raffinose synthase and stachyose synthase generates tetrasaccharides (stachyose) from the reaction of raffinose and galactionol (Peterbauer and Richter, 2001). Some stachyose synthases have an ability to add two galactoses to raffinose, producing the pentasaccharides verbascose.

Myo-inositol 3-phosphate synthase (MIPS)

The synthesis of myo-inositol is from glucose 6-phosphate and D-myo-inositol 3-phosphate. The enzyme that converts glucose 6-phosphate to D-myo-inositol 3-phosphate is D-myo-inositol 3-phosphate synthase (MIPS). MIPS is a highly conserved enzyme that has been identified throughout all biological kingdoms (Majumder et al. 1997; Bachhawat and Mande 2000). In the soybean genome, there are four *MIPS* gene. Of these genes, only *MIPS1* is highly expressed in dry seeds (Hegeman et al., 2001; Hitz et al., 2002; Nunes et al., 2006; Chiera and Grabau, 2007). A soybean *MIPS1* mutant line LR33 (Kerr and Sebastian, 2000) contains a 50 % decrease in phytic acid, raffinose, stachyose and an increase in sucrose. It is the result of a single base pair polymorphism in a conserved region of the *MIPS1* gene (Hitz et al., 2002). Two novel soybean mutations for low phytic acid phenotype were isolated. A mutant allele of one of soybean lines contains 2 bp deletion in *MIPS1* (Yuan et al., 2007). Another mutation was mapped on LG B2 (Chromosome 14), linked with SSR markers such as Satt 416 and Satt 168 (Yuan et al., 2007). Based on a candidate gene sequencing with low phytate soybean lines CX1834 produced by chemical mutagenesis, two multidrug resistance associated protein

encoding genes (MRPs) were identified and found to contain mutations that reduced phytic acid accumulation in seeds (Maroof et al., 2009; Gillman et al., 2009).

Galactinol synthases

Galactinol synthase plays a key role in the carbon partitioning between sucrose and RFO (Saravitz et al., 1987). Overexpression of a galactinol synthase gene in transgenic Arabidopsis plants results in an increase in the levels of endogenous galactinol and raffinose under normal conditions (Taji et al., 2002). Galatinol synthase activity is detected in seed development in plants. During soybean seed development, the desiccation stage of the seed is accompanied by an increase in RFO and galactinol synthase activity (Saravitz et al., 1987). Also, the enzymatic activity of galactinol synthase is positively associated with the concentration of stachyose. Additionally, raffinose concentrations are linearly related to galactinol formation by galactinol synthase (Saravotz et al., 1987)

Carbohydrate profile in soybean

A search of the GenBank database revealed that raffinose synthase sequences have been annotated for pea (AJ426475), cucumber (AF073744), maize (Q575Z6, Q575Z7, Q575Z8), rice (BAD321.1), grape (AM430487), arabidopsis (AC007138) and soybean (EU651888).

The soybean plant introduction PI 200508 was shown to have lower RFO and an increased sucrose concentration compared to wild-type plants (Kerr and Sebastian, 2002). In addition, another study indicated that a mutant line with reduced RFO has increased levels of

sucrose (Hitz et al., 2002). There were no negative agronomic characteristics associated with soybean line PI 200508 (Neus et al., 2005). The genetic basis of the reduced RFO trait derived from PI 200508 was reported. They found that a novel allele contained a deletion of three base pairs within the raffinose synthase 2 coding sequence (RS2, Glyma06g18890), resulting in a deletion in a highly conserved region of a tryptophan residue at position 331 (*rs2W331-*) (Dierking and Bilyeu, 2008). A mutation of this region in soybean associates with the increased sucrose level and the decreased raffinose and stachyose phenotype. Targeting Induced Local Lesions IN Genomes (TILLING) is a reverse genetics technique that is a high throughput method to identify mutations within target genes (Colbert et al., 2001; McCallum et al., 2000). By TILLING, another mutation in the RS2 gene was identified in soybean. This mutant allele of the RS2 gene contains a missense mutation resulting in the change of isoleucine from threonine at amino acid position 107 (*rs2T107I*) (Dierking and Bilyeu, 2009). Soybean lines with the *rs2T107I* alleles also showed reductions in raffinose and stachyose along with an increase in sucrose. However, soybean lines with the *rs2W331-* allele had a lower level of RFO than lines with the *rs2T107I* allele. Soybean lines that carried the *rs2T107I* allele showed a low RFO phenotype (Hagely et al., 2013). Genotypic and phenotypic investigations have revealed that lines containing the low RFO (*rs2W331-*) allele and a functional RS3 allele consistently have a low RFO phenotype, while lines containing the mutant *rs2W331-* allele with variant alleles of raffinose synthase 3 gene (*rs3SNP5*) have the ultra-low RFO phenotype, characterized by RFO levels below approximately 1 % of the seed (Hagely, 2013).

Soybean lines that had different genotypes with respect to RS2 and RS3 genes were tested for RFO in only one location in a single environment (Hagely et al., 2013). This study

indicated that the carbohydrate profile of soybean was somewhat variable between genotypes. However, it can be distinguishable from the carbohydrate profile of seeds from each genotypic classes of RS2 and RS3 (Hagely et al., 2013). Another group had reported the carbohydrate profile of seven soybean accessions grown at three very different geographic locations (Kumar et al., 2010). The results demonstrated that the sucrose content increased with cooler growing conditions. In addition, another study was conducted with soybean genotypes containing different combinations of RS2 and RS3 alleles tested in five different planting dates (Bilyeu and Wiebold, 2016). They demonstrated that soybean lines in late planting date have more content of sucrose than ones in an early planting date. They suggested that higher sucrose contents were able to be produced when there were cooler temperatures during the pod filling period (Bilyeu and Wiebold, 2016; Kumar et al., 2010). However, raffinose and stachyose values were variable (Kumar et al., 2010). It is necessary to further study the role of the environment in affecting the stability in the carbohydrate profile in soybean seed throughout U.S. production environments.

Association with RFO content and soybean maturity

RFO accumulate during the process of cold acclimation (Gilmour et al., 2000; Koster et al., 1992). In leaves of tomato and Arabidopsis seedlings, raffinose synthase increased under environmental stress such as cold (Downie et al., 2003; Zuther et al., 2004). Thus, accumulation of RFO is highly associated with temperature. Wolf et al. (1982) studied the effect of low and high temperatures on carbohydrates. They concluded that raffinose levels in soybean seed were unaffected by low and high temperature, whereas sucrose and stachyose decreased at the highest temperature (Wolf et al., 1982). Considerable genetic variation for the content of RFO of mature

seeds has been reported (Pattee et al., 2000). Bellaloui et al. (2010) reported the effect of genotypic background of a maturity gene (E gene) on sugars. They indicated that there is no significant relationship with maturity and raffinose but, there was a negative correlation between maturity and sucrose as well as between stachyose and total carbohydrate (Bellaloui et al., 2010).

Soybean lines from four different genotypic classes were tested for carbohydrate profile on two different planting dates (Hagely et al., 2013). The results indicated that there was no consistent impact of maturity date on carbohydrate profile, but distinct genotypic classes were overall distinguishable (Hagely et al., 2013). Another group measured the carbohydrate profiles of seven soybean accessions at three different geographical locations. They found that sucrose was significantly higher at cooler locations, while raffinose and stachyose were genotype dependent at the growing locations (Kumar et al., 2010). They suggested that soybean accessions grown at cooler locations may be better able to produce soy food due to the carbohydrate characteristics (Kumar et al., 2010). Also, another study was conducted with soybean lines containing altered raffinose synthase 2 alleles in five distinct locations in addition to five successive planting dates. They demonstrated that cooler temperatures during seed development were highly correlated with increased sucrose and decreased stachyose (Bilyeu and Wiebold, 2016).

Quantitative trait loci of sucrose content in soybean

Sucrose is a primary constituent of soybean seed. Sucrose content is closely associated with the quality and taste of many soyfood products. An understanding of the genetic factors that contribute to sucrose content in soybean could help breeders accelerate the development of

varieties for food quality purposes. Maughan et al (2000) analyzed 149 F_2 individuals from a cross of *G. Max* and *G. soja* with 178 polymorphic genetic markers, including RFLPs, SSRs, and RAPDs. They concluded that seventeen marker loci, mapping to seven different genomic regions (Chromosome 5, 7, 8, 13, 15, 19 and 20), were significantly associated with sucrose variation (Maughan et al., 2000). Also, they reported that a comparison of mapping results for sucrose and those for protein and oil in soybean suggested that seed composition traits were inherited as clusters of linked loci (Maughan et al., 2000). Another group also tested 115 F_{10} lines developed by single seed descent (SSD) for molecular markers to identify the genomic regions significantly associated with sucrose content (Kim et al., 2005). They found that QTLs for sucrose content are located at chromosome 11 (satt197), chromosome 2 (satt546), and chromosome 19 (satt523 and satt278) (Kim et al., 2005). Two of them (chromosome 2 and 19) were associated with oligosaccharide content. Two flanking markers on chromosome 19 tagged the major QTL for sucrose, which showed 21.4 % phenotypic variation for sucrose. Another RIL population was used to map QTLs on chromosome 12 and 16 for sucrose content (Kim et al., 2006). For oligosaccharide contents, four QTLs (chromosome 6, 12, 16 and 19) were detected. Phenotypic variations for each tagging markers were less than 10 % for sucrose content. Although the two RIL populations in the two studies shared the high sucrose donor soybean line "Keunolkong", the QTLs for sucrose in the two mapping populations did not overlap. Zeng et al. (2014) have recently reported on sucrose content in PI 243545 which is a germplasm introduced from Japan (USDA-ARS, 2009). Three QTLs were mapped on chromosome 5, 9 and 16 which have 46 %, 10 % and 8 % of phenotypic variation for sucrose content.

Soybean germplasm collection

The USDA began recording introduced soybean lines, generally referred to as plant introductions, in 1898 and assigning each a permanent Plant Introduction (PI) number. Now, more than 21,000 accessions are found in the USDA soybean germplasm collection. Information on germplasm can be found in the USDA's database (http://www.grin-global.org) (Perry and McIntosh, 1991). All of the accessions within the U.S. soybean germplasm collection are freely available to scientists worldwide. These soybean lines are an important source for new genes for disease resistance, seed composition modification, tolerance to various abiotic stresses and, in recent years, the source of new genes for high yield.

The U.S. soybean germplasm is maintained at the USDA Soybean Germplasm Collection, Champaign/Urbana, Illinois. The. soybean germplasm collection currently has 21,729 accessions including soybeans, wild annual soybeans and wild perennial *Glycine* species, including 19,541 *Glycine max*, 1,181 *Glycine soja*, and 19 perennial species with 1,007 lines (Hymowitz, 2008).

Molecular Breeding

Marker-Assisted Selection (MAS) using DNA markers in plants is a tool to select plants containing specific agronomic traits (Collard and Mackill, 2008). The markers universally used in plants are Simple Sequence Repeats (SSRs) or Single Nucleotide Polymorphisms (SNPs) (Choi *et al*., 2007; Gupta *et al*., 2008; Rafalski, 2002; Song *et al*., 2004). Molecular breeding is becoming very useful and has made remarkable progress in crop improvement. With the help of molecular technologies, soybean scientists can easily select locations of genes and markers tightly linked to desirable traits of interest. MAS saves time and labor in plant breeding programs

(Koebner and Summers, 2003). To use MAS, genetic markers must be identified as candidate genes or mapped to tag useful genes for easier selection of economical traits. Populations that are segregating for phenotype(s) in question are used (Tanksley, 1993) to locate and tag the gene using SSR and SNP markers. The DNA markers for specific genes to be employed in MAS are determined by statistical methods used in mapping procedures (Kearsey, 1998). MAS is less effective for quantitative traits controlled by many genes, whereas major genes or QTLs for quantitative traits can have a strong correlation with the phenotypic variation (Jiang, 2013). In addition, the markers which are tightly linked to interested genes can be used for MAS in plant breeding programs. However, it is possible to have a chance to break the region between candidate gene and linked marker through genetic recombination. With the identification of the molecular basis for a specific trait through candidate gene methods, the causative mutant allele is identified and a "perfect" molecular marker for a trait can be developed to use in breeding program. Unlike linked markers, perfect markers are by definition not subject to recombination between the marker and the trait.

Nowadays, genome sequencing has been completed for many plants and animals including soybeans. The genome sequencings for soybean and wild soybean were completed (Kim *et al*., 2010; Schmutz *et al*., 2010). Scientists can use genome-sequencing information of soybean to find genes related to traits of interest in order to develop new soybean cultivars. In addition, a universal linkage panel including 1,536 SNPs was developed to use for mapping studies (Hyten *et al*., 2010) and more recently, a BeadChip containing more than 50,000 SNPs was developed to genotype a high-density genetic map of soybean (Song *et al*., 2013). Song et al. (2015) recently have used this SoySNP50K BeadChip with the entire soybean germplasm

collection that includes 18,480 domesticated soybean and 1168 wild soybean accessions. Lee et al. (2015) recently have developed an 180K SNP genotyping array for mapping. Additionally, the genotyping by sequencing (GBS) in soybean was recently developed and utilized. The GBS method was the new genotyping approach with low cost and high resolution (Sonah et al., 2013).

A segregating plant population is usually used for development of mapping populations. Parental lines are homozygous; one contains the trait of interest and the other does not have the trait. Several populations are used for mapping the gene in plants. These populations consist of recombinant inbred lines (RILs), backcross (BC) and doubled haploid (DH) (Collard et al., 2005). Inbreeding from individual F_2 plants allows the construction of RILs, containing a unique combination of chromosomal segments from the original parents. As six to eight generations are required, producing RILs is a major disadvantage. To overcome this disadvantage, a winter nursery can be used to advance breeding populations for multiple generations in a single year. However, as individual RILs are homozygous and can be multiplied, it means that they can be used in replicated trials across different locations and years (Collard et al., 2005). BC populations are derived by crossing the F_1 hybrid to one of the parents. The purpose of BC is to recover the genotype of one of parental line with substitution of a desirable allele. Either repeated selfing or backcrossing produces a mapping population designated as near isogenic lines. DH populations may be produced by the induction of chromosome doubling from anthers of F_1 plants, if the plant species are available for tissue culture. The major advantage of DH populations is that they produce true-breeding lines that can be reproduced like RILs (Collard et al., 2005).

The purpose of this project is to understand and characterize the variation in sucrose and raffinose family of oligosaccharides for human and animals in soybean seed. First, we have surveyed the genomic region containing the raffinose synthase 3 gene and identified an improved SNP position that associates with the ultra-low RFO phenotype in conjunction with mutant RS2 alleles. Secondly, an environmental stability study in two locations over two years was conducted. It is necessary to study the role of the environment in affecting the stability in the carbohydrate profile in soybean seed with variant alleles of RS2 and RS3 throughout the soybean production areas in the U.S. Thus, maturity gene combinations targeted for adaptation of a specific region were fixed for the development of soybean lines with different RS2 and RS3 alleles and those soybean lines were grown in different maturity group regions. This study produced one year of results for the carbohydrate profile of soybean seed to develop soybean cultivars with stable carbohydrate profile trait in different maturity groups for breeders. Lastly, genotyping by sequencing approach was used to map the QTLs for sucrose content from parental lines having identical maturity gene combinations but contrasting for high sucrose content with normal RFO and high sucrose content with ultra-low RFO. Individuals in the mapping population matured in similar time in order to reduce the environmental effect on the accumulation of sucrose content.

References

Abe, J., D. Xu, A. Miyano, K. Komatsu, A. Kanazawa, and Y. Shimamoto. (2003). Photoperiod-insensitive Japanese soybean landraces diff er at two maturity loci. Crop Sci. 43:1300–1304.

Anderson-Hafermann, J. C., Zhang, Y., Parsons, C. M., and Hymowitz, T. (1992). Effect of heating on nutritional quality of conventional and Kunitz trypsin inhibitor-free soybeans. Poultry science, 71(10), 1700-1709.

Alscher R.G., Donahue JL, Cramer C. (1997) Reactive oxygen species and antioxidants: relationships in green cells. Physiol Plant 100: 224-233

Amuti, K.S. and Pollard, C.J. (1977). Soluble carbohydrates of dry and developing seeds. Phytochemistry 16, 5, (May 1977) 529-532, ISSN 0031-9422

Bachhawat, N., and S. C. Mande, (2000). Complex evolution of theinositol-1-phosphate synthase gene among archaea and eubacteria. Trends Genet. 16, 111—113.

Bach Knudsen KE, Aman P, Eggum BO (1987). Nutritive value of danish-grown barley varieties, I, carbohydrates and other major constituents. J Cereal Sci 6:173-186

Bailly, C., Audigier C., Ladonne F., Wagner M.H., Coste F., Corbineau F., and Come D. (2001) Changes in oligosaccharide content and antioxidant enzyme activities in developing bean seeds as related to acquisition of drying tolerance and seed quality. Journal of Experimental Botany 52:701-708.

Beebe D.U. Turgeon R. (1992) Localization of galactinol, raffinose, and stachyose synthesis in Cucurbita pepo leaves. Planta 188:354-361.

Bewley, J.D., and M. Black. (1994) Seeds: Physiology of Development and Germination. Second ed. Plenum Press, New York.

Blöchl A., Peterbauer T., and Richter A. (2007). Inhibition of raffinose oligosaccharide breakdown delays germination of pea seeds. Journal of Plant Physiology 164:1093-1096.

Bowler C., Montagu M.V., Inze D. (1992) Superoxide dismutase and stress tolerance. Annu Rev Plant Physiol Plant Mol Biol 43: 83–116

Bernard, R.L. (1971). Two major genes for time of flowering and maturity in soybeans. Crop Sci. 11:242–244.

Buzzell, R.I. (1971). Inheritance of a soybean flowering response to fluorescent-daylength conditions. Can. J. Genet. Cytol. 13:703–707.

Buzzell, R.I., and H.D. Voldeng. (1980). Inheritance of insensitivity to long daylength. Soybean Genet. Newsl. 7:26–29.

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.

Brummer EC, Graef GL, Off J, Wilcox JR, Shoemaker RC (1997) Mapping QTL for seed protein and oil content in eight soybean populations. Crop Sci 37:370-378

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 () Lines. Crop science, 50(5), 1978-1987.

Bonato, E.R., and N.A. Vello. (1999). E6, a dominant gene conditioning early flowering and maturity in soybeans. Genet. Mol. Biol. 22:229–232.

Colbert, T., B.J. Till, R. Tompa, S. Reynolds, M.N. Steine, A.T. Yeung, C.M. McCallum, L. Comai, and S. Henikoff. (2001). High-throughput screening for induced point mutations. Plant Physiology 126:480-484.

Cober, E.R., Molnar, S.J., Charette, M. and Voldeng, H.D., (2010). A new locus for early maturity in soybean. Crop Science, 50(2), pp.524-527.

Castillo, E. M., De Lumen, B. O., Reyes, P. S., and De Lumen, H. Z. (1990). Raffinose synthase and galactinol synthase in developing seeds and leaves of legumes. Journal of Agricultural and Food Chemistry, 38(2), 351-355.

Cober, E.R., and H.D. Voldeng. (2001). A new soybean maturity and photoperiod-sensitivity locus linked to E1 and T. Crop Sci. 41:698–701.

Crookston R. K. and D. S. Hill. (1978). A visual indicator of the physiological maturity of soybean seed. Crop sciences 18:867 870

Correll, D.L. (1998). Th e role of phosphorus in the eutrophication of receiving waters: A review. J. Environ. Qual. 27:261–266.

Collard, B. C. Y., Jahufer, M. Z. Z., Brouwer, J. B., and Pang, E. C. K. (2005). An introduction to markers, quantitative trait loci (QTL) mapping and marker-assisted selection for crop improvement: the basic concepts. Euphytica, 142(1-2), 169-196.

Collard B.C., Mackill D.J. (2008) Marker-assisted selection: an approach for precision plant breeding in the twenty-first century. Philosophical Transactions of the Royal Society B: Biological Sci. 363:557-572.

Chiera, J.M. and Grabau, E.A., (2007). Localization of myo-inositol phosphate synthase (GmMIPS-1) during the early stages of soybean seed development.Journal of experimental botany, 58(8), pp.2261-2268.

Choi I.-Y., Hyten D.L., Matukumalli L.K., Song Q., Chaky J.M., Quigley C.V., Chase K., Lark K.G., Reiter R.S., Yoon M.-S. (2007) A soybean transcript map: gene distribution, haplotype and single-nucleotide polymorphism analysis. Genetics 176:685-696.

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.

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

Dierking, E.C. and Bilyeu, K.D. (2009). Raffinose and stachyose metabolism are not required for efficient soybean seed germination. Journal of Plant Physiology 166(12): 1329-1335

Downie B., Gurusinghe S., Dahal P., Thacker R.R., Snyder J.C., Nonogaki H., Yim K., Fukanaga K., Alvarado V., Bradford K.J. (2003) Expression of a GALACTINOL SYNTHASE gene in tomato seeds is up-regulated before maturation desiccation and again after imbibition whenever radicle protrusion is prevented. Plant Physiol 131: 1347–1359

DiPietro, C. M., and Liener, I. E. (1989). Heat inactivation of the Kunitz and Bowman-Birk soybean protease inhibitors. Journal of agricultural and food chemistry, 37(1), 39-44.

Franklin KA, Quail PH (2010) Phytochrome functions in Arabidopsis development. J Exp Bot 61:11–24.

Fehr, W.R. (2007) Breeding for modified fatty acid composition in soybean. Crop Science 47:S-72-87.

Gerde JA, White PJ (2008) Lipids. In: Johnson LA, White PJ, Galloway R (eds) Soybeans: chemistry, production, processing, and utilization. AOCS Press, Urbana, IL, pp 193-227

Gupta P., Varshney R.K., Sharma P., Ramesh B. (2008) Molecular markers and their applications in wheat breeding. Plant Breeding 118:369-390.

Gillman, J.D., Pantalone, V.R. and Bilyeu, K., (2009). The low phytic acid phenotype in soybean line CX1834 is due to mutations in two homologs of the maize low phytic acid gene. The Plant Genome, 2(2), pp.179-190.

Gilmour, S. J., Sebolt, A. M., Salazar, M. P., Everard, J. D., and Thomashow, M. F. (2000). Overexpression of the Arabidopsis CBF3transcriptional activator mimics multiple biochemical changes associated with cold acclimation. Plant physiology, 124(4), 1854-1865.

Gitzelmann, R., and Auricchio, S. (1965). The handling of soya alpha-galactosides by a normal and a galactosemic child. Pediatrics, 36(2), 231-235.

Hegeman, C.E., Good, L.L. and Grabau, E.A., (2001). Expression of D-myo-inositol-3-phosphate synthase in soybean. Implications for phytic acid biosynthesis. Plant Physiology, 125(4), pp.1941-1948.

Heaney, R.P., Weaver, C.M. and Fitzsimmons, M.L., (1991). Soybean phytate content: effect on calcium absorption. The American journal of clinical nutrition, 53(3), pp.745-747.

Hitz, W. D., Carlson T. J., Kerr P. S., and Sebastian S. A., (2002): Biochemical and molecular characterization of a mutation that confers a decreased raffinosaccharide and phytic acid phenotype on soybean seeds. Plant. Physiol. 128, 650—660.

Hagely, K.B., (2013). Improving soybean seed composition through molecular breeding for Raffinose family oligosaccharides, lectin, and trypsin inhibitors(Doctoral dissertation, University of Missouri).

Hagely, K. B., Palmquist, D., and Bilyeu, K. D. (2013). Classification of distinct seed carbohydrate profiles in soybean. Journal of agricultural and food chemistry, 61(5), 1105-1111.

Hymowitz T (1970) On the domestication of the soybean. Econ Bot 24:408-421

Hymowitz T (2008) The history of the soybean. In: Johnson LA, White PJ, Galloway R (eds) Soybeans: chemistry, production, processing, and utilization. AOCS Press, Urbana, IL, pp 1-31

Hymowitz, T. (1990) Soybeans: The Success Story, p. 159-163, In J. Janick and J. E. Simon, eds. Advances in new crops. Timber Press, Portland, OR.

Hymowitz, T., and W.R. Shurtleff. (2005) Debunking Soybean Myths and Legends in the Historical and Popular Literature. Crop Science 45:473-476.

Hymowitz T, Harlan JR (1983) The introduction of the soybean to North America by Samuel Bowen in 1765. Econ Bot 37:371-379

Hymowitz T, Newell CA (1980) Taxonomy, speciation, domestication, dissemination, germplasm resources, and variation in the genus Glycine. In: Summerfield RJ, Bunting

AH (eds) Advances in legume science. Royal Botanical Garden, Kew, England, pp 251-264

Hsu, S.H., Hadley, H.H. and Hymowitz, T. (1973). Changes in carbohydrate contents of germinating soybean seeds. Crop Science 13:407-410

Hyten D.L., Choi I.-Y., Song Q., Specht J.E., Carter T.E., Shoemaker R.C., Hwang E.-Y., Matukumalli L.K., Cregan P.B. (2010) A high density integrated genetic linkage map of soybean and the development of a 1536 universal soy linkage panel for quantitative trait locus mapping. Crop Sci. 50:960-968.

Hincha, D. K. (1990). Differential effects of galactose containing saccharides on mechanical freeze-thaw damage to isolated thylakoid membranes. Cryo letters.

Jiang, G.L., (2013). Molecular markers and marker-assisted breeding in plants. Plant Breeding from Laboratories to Fields, pp.45-83.

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(1), 106-112.

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(5), pp.533-538.

Kim M.Y., Lee S., Van K., Kim T.-H., Jeong S.-C., Choi I.-Y., Kim D.-S., Lee Y.-S., Park D., Ma J. (2010) Whole-genome sequencing and intensive analysis of the undomesticated soybean (Glycine soja Sieb. and Zucc.) genome. Proceedings of the National Academy of Sciences 107:22032-22037.

Koster, K. L., Lynch, D. V. (1992). Solute accumulation and compartmentation during the cold acclimation of Puma rye. Plant Physiology, 98(1): 108-113.

Koster, K.L., Leopold, A.C. (1988). Sugars and desiccation tolerance in seeds. Plant Physiology 829-832

Koebner R., Summers R.W. (2003) 21st century wheat breeding: plot selection or plate detection? Trends in Biotechnology 21:59-63.

Kearsey M. (1998) The principles of QTL analysis (a minimal mathematics approach). J Exp Bot 49:1619-1623.

Karr-Lilienthal L.K., Grieshop C.M., Spears J.K., Fahey Jr. G.C. (2005) Amino acid, carbohydrate, and fat composition of soybean meals prepared at 55 commercial U.S. soybean processing plants. J Agric Food Chem 53:2146-2150

Kwanyuen, P., Pantalone, V. R., Burton, J. W., and Wilson, R. F. (1997). A new approach to genetic alteration of soybean protein composition and quality. Journal of the American Oil Chemists' Society, 74(8), 983-987.

Kerr, P. S. and Sebastian S. A. (2000). Soybean products with improved carbohydrate composition and soybean plants. United States, E. I. du Pont de Nemours and Company. 6,147,193.

Kumar, V. Rani, A.; Goyal, L.; Dixit, A. K.; Manjaya, J. G.; Dev, J.; Swamy, M. (2010) Sucrose and raffinose family oligosaccharides (RFOs) in soybean seeds as influenced by genotype and growing location. J. Agric. Food Chem., 58 (8), 5081−5085.

Lee, Y.G., Jeong, N., Kim, J.H., Lee, K., Kim, K.H., Pirani, A., Ha, B.K., Kang, S.T., Park, B.S., Moon, J.K. and Kim, N., (2015). Development, validation and genetic analysis of a large soybean SNP genotyping array. The Plant Journal,81(4), pp.625-636.

Lynch, S.R., Dassenko, S.A., Cook, J.D., Juillerat, M.A. and Hurrell, R.F., (1994). Inhibitory effect of a soybean-protein--related moiety on iron absorption in humans. The American journal of clinical nutrition, 60(4), pp.567-572.

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

Liu B, Kanazawa A, Matsumura H, Takahashi R, Harada K, et al. (2008) Genetic redundancy in soybean photoresponses associated with duplication of the phytochrome A gene. Genetics 180: 995–1007.

Lowell, C.A. and Kuo, T.M. (1989). Oligosaccharide metabolism and accumulation in developing soybean seeds. Crop Science 29, 2

Langewisch, T., Zhang, H., Vincent, R., Joshi, T., Xu, D. and Bilyeu, K., (2014). Major soybean maturity gene haplotypes revealed by SNPViz analysis of 72 sequenced soybean genomes. PloS one, 9(4), p.e94150.

Maughan, P. J., Maroof, M. S., and Buss, G. R. (2000). Identification of quantitative trait loci controlling sucrose content in soybean (Glycine max). Molecular Breeding, 6(1), 105-111.

McBlain, B.A., and R.L. Bernard. (1987). A new gene affecting the time of flowering and maturity in soybeans. J. Hered. 78:160–162.

Middelbos I.S., Fahey Jr. G.C. (2008) Soybean carbohydrates. In: Johnson LA, White PJ, Galloway R (eds) Soybeans: chemistry, production, processing, and utilization. AOCS Press, Urbana, IL, pp 269-296

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

McCallum, C.M., L. Comai, E.A. Greene, and S. Henikoff. (2000). Targeting induced local lesions IN genomes (TILLING) for plant functional genomics. Plant Physiology 123:439-442

Maroof, M.A., Glover, N.M., Biyashev, R.M., Buss, G.R. and Grabau, E.A., (2009). Genetic basis of the low-phytate trait in the soybean line CX1834. Crop science, 49(1), pp.69-76.

Majumder, A. L., Johnson M. D., and Henry S. A., (1997) 1L-myoinositol- 1-phosphate synthase. Biochim. Biophys. Acta. 1348, 245—256.

Molnar, S. J., Rai, S., Charette, M., and Cober, E. R. (2003). Simple sequence repeat (SSR) markers linked to E1, E3, E4, and E7 maturity genes in soybean. Genome, 46(6), 1024-1036.

Nishizawa A., Yabuta Y.,Shigeoka S. (2008) Galactinol and raffinose constitute a novel function to protect plants from oxidative damage. Plant Physiology 147: 1251–1263.

Nishizawa A., Yabuta Y., Yoshida E., Maruta T., Yoshimura K., Shigeoka S. (2006) Arabidopsis heat shock transcription factor A2 as a key regulator in response to several types of environmental stress. Plant J 48: 535–547

Nunes, A.C., Vianna, G.R., Cuneo, F., Amaya-Farfán, J., de Capdeville, G., Rech, E.L. and Aragão, F.J., (2006). RNAi-mediated silencing of the myo-inositol-1-phosphate synthase

gene (GmMIPS1) in transgenic soybean inhibited seed development and reduced phytate content. Planta, 224(1), pp.125-132.

Neus, J.D., W.R. Fehr, and S.R. Schnebly. (2005). Agronomic and seed characteristics of soybean with reduced raffinose and stachyose. Crop Science 45:589-592.

Orf, J. H., and Hymowitz, T. (1979). Inheritance of the absence of the Kunitz trypsin inhibitor in seed protein of soybeans. Crop Science, 19(1), 107-109.

Obendorf, R.L. (1997) Oligosaccharides and galactosyl cyclitols in seed desiccation tolerance. Seed Science Research 7:63-74.

Obendorf, R.L., Ashworth, E.N. and Rytko, G.T. (1980). Influence of seed maturation on germinability in soybean. Crop Science 20, 4

Obendorf, R.L., Zimmerman, A.D., Zhang, Q., Castillo, A., Kosina, S.M., Bryant, E.G., Sensenig, E.M., Wu, J. and Schnebly, S.R. (2009). Accumulation of soluble carbohydrates during seed development and maturation of low-raffinose, lowstachyose soybean. Crop Science 49, 1

Obendorf, R. L., A. D. Zimmerman, P. A. Ortiz, A. G. Taylor and S. R. Schnebly (2008). Imbibitional chilling sensitivity and soluble carbohydrate composition of low raffinose, low stachyose soybean seed. Crop Sci. 48(6): 2396---2403.

Peterbauer, T., Lahuta, L. B., Blöchl, A., Mucha, J., Jones, D. A., Hedley, C. L., and Richter, A. (2001). Analysis of the raffinose family oligosaccharide pathway in pea seeds with contrasting carbohydrate composition. Plant Physiology, 127(4), 1764-1772.

Peterbauer, T. and Richter, A. (2001). Biochemistry and physiology of raffinose family oligosaccharides and galactosyl cyclitols in seeds. Seed Science Research, 11(3), 185-198.

Price, K.R., J. Lewis, G.M. Wyatt, and G.R. Fenwick. (1988). Flatulence—causes, relation to diet and emedies. Nahrung 32:609–623.

Pharr, D. M., and Sox, H. N. (1984). Changes in carbohydrate and enzyme levels during the sink to source transition of leaves of< i> Cucumis sativus</i> L., a stachyose translocator. Plant science letters, 35(3), 187-193.

Parsons, C.M., Y. Zhang, and M. Araba. (2000). Nutritional evaluation of soybean meals varying in oligosaccharide content. Poult. Sci. 79:1127–1131.

Pennycooke, J. C., Jones, M. L., and Stushnoff, C. (2003). Down-regulating α-galactosidase enhances freezing tolerance in transgenic petunia. Plant Physiology, 133(2), 901-909.

Perry, M. C., and McIntosh, M. S. (1991). Geographical patterns of variation in the USDA soybean germplasm collection: I. Morphological traits. Crop Science, 31(5), 1350-1355.

Pattee, H.E., T.G. Isleib, F.G. Giesbrecht, and R.F. McFeeters. (2000) Investigations into genotypic variations of peanut carbohydrates. J. Agric. Food Chem. 48:750–756.

Raboy, V. (2002). Progress in breeding low phytate crops. J. Nutr. 132:503S–505S.

Raboy, V. (2007a) Seed phosphorus and the development of low-phytate crops. In: B.L.

Raboy, V., Dickinson, D.B. and Below, F.E., (1984). Variation in seed total phosphorus, phytic acid, zinc, calcium, magnesium, and protein among lines of Glycine max and G. soja. Crop Science, 24(3), pp.431-434.

Raboy, V. (2007b). The ABCs of low-phytate crops. Nat. Biotechnol. 25:874–875. doi:10.1038/nbt0807-874

Rafalski J.A. (2002) Novel genetic mapping tools in plants: SNPs and LD-based approaches. Plant Sci 162:329-333.

Ruttloff , H., A. Täufel, W. Krause, H. Haenel, and K. Täufel. (1967). The intestinal enzymatic decompositon of galacto-oligosaccharides in the human and animal intestine, with particular regard to Lactobacillus bifi dus. Part II: On the intestinal behaviour of lactulose. Nahrung 11:39–46.

Singh, R. J., and Hymowitz, T. (1999). Soybean genetic resources and crop improvement. Genome, 42(4), 605-616.

SoyStats (2014) Available at http://www.soystats.com.

Soybase. 2014. A USDA–ARS Plant Genome Program soybean database. Available at http://soybase.org.

Soyfoods Descriptions [Online] http://soyfoods.com/soyfoodsdescriptions/descriptions.html

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

Suarez, F.L., J. Springfi eld, J.K. Furne, T.T. Lohrmann, P.S. Kerr, and M.D. Levitt. (1999). Gas production in humans ingesting a soybean fl our derived from beans naturally low in oligosaccharides. Am. J. Clin. Nutr. 69:135–139.

Selvendran RR, Stevens BJH, Du Pont MS (1987) Dietary fiber: chemistry, analysis, and properties. Adv Food Res 31:117-209

Schillinger, J.A., Dierking, E.C. and Bilyeu, K.D., Schillinger Genetics, Inc. and The United States Of America As Represented By The Secretary Of Agriculture, (2013). Soybeans having high germination rates and ultra-low raffinose and stachyose content. U.S. Patent 8,471,107.

Shigeoka S., Ishikawa T., Tamoi M., Miyagawa Y., Takeda T., Yabuta Y., Yoshimura K. (2002) Regulation and function of ascorbate peroxidase isoenzymes. J Exp Bot 53: 1305–1319

Senser, M., and Kandler, O. (1967). Galactinol, ein Galactosyldonor für die Biosynthese der Zucker der Raffinosefamilie in Blättern. Z. Pflanzenphysiol, 57, 376-388.

Song Q., Marek L., Shoemaker R., Lark K., Concibido V., Delannay X., Specht J.E., Cregan P. (2004) A new integrated genetic linkage map of the soybean. TAG Theoretical and Applied Genetics 109:122-128.

Song Q., Hyten D.L., Jia G., Quigley C.V., Fickus E.W., Nelson R.L., Cregan P.B. (2013) Development and evaluation of SoySNP50K, a high-density genotyping array for soybean. Plos one 8:e54985.

Song, Q., Hyten, D.L., Jia, G., Quigley, C.V., Fickus, E.W., Nelson, R.L. and Cregan, P.B., (2015). Fingerprinting soybean germplasm and its utility in genomic research. G3: Genes| Genomes| Genetics, 5(10), pp.1999-2006.

Schmutz J., Cannon S.B., Schlueter J., Ma J., Mitros T., Nelson W., Hyten D.L., Song Q., Thelen J.J., Cheng J. (2010) Genome sequence of the palaeopolyploid soybean. Nature 463:178-183.

Sonah H, Bastien M, Iquira E, Tardivel A, Légaré G, Boyle B (2013) An Improved enotyping by Sequencing (GBS) Approach Offering Increased Versatility and Efficiency of SNP Discovery and Genotyping. PloS One. 2013;8(1):e54603.

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.

Tanksley S.D. (1993) Mapping polygenes. Annual Rev. Genetics 27:205-233.

Tarczynski M.C., Jensen R.G., Bohnert H.J. (1993) Stress protection of transgenic tobacco by production of the osmolyte mannitol. Science, 259, 508–510.

Tsubokura Y, Matsumura H, Xu M, Liu B, Nakajima H, et al. (2013) Genetic variation in soybean at maturity locus E4 is involved in adaptation to long days at high latitudes. Agronomy 3: 117–134.

Turgeon, R. (1996) Phloem loading and plasmodesmata. Trends Plant Sci., 1, 418 423

Taji, T., Ohsumi, C., Iuchi, S., Seki, M., Kasuga, M., Kobayashi, M., Shinozaki, K. (2002). Important roles of drought-and cold-inducible genes for galactinol synthase in stress tolerance in Arabidopsis thaliana. The Plant Journal, 29(4), 417-426.

Verma, D. P. S., and Shoemaker, R. C. (1996). Soybean: genetics, molecular biology and biotechnology. CAB international.

Watanabe, T. (1997). Science of tofu. Kyoto, Japan: Food Journal Co, 14-54.

Watanabe, S., Xia, Z., Hideshima, R., Tsubokura, Y., Sato, S., Yamanaka, N., Takahashi, R., Anai, T., Tabata, S., Kitamura, K. and Harada, K., (2011). A map-based cloning strategy employing a residual heterozygous line reveals that the GIGANTEA gene is involved in soybean maturity and flowering. Genetics, 188(2), pp.395-407.

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

Wilcox, J. R. (1987). Soybeans: improvement, production and uses (No. Ed. 2). American Society of Agronomy, Crop Science Society of America, Soil Science Society of America.

Wilson, R. F. (1987). Seed metabolism. Agronomy (USA).

Wilson RF (2004) Seed composition. In: Boerma HR, Specht JE (eds) Soybeans: improvement, production and uses. ASA, CSSA and SSSA, Madison, WI, pp 621-677

Wilson, R. P., and Poe, W. E. (1985). Effects of feeding soybean meal with varying trypsin inhibitor activities on growth of fingerling channel catfish. Aquaculture, 46(1), 19-25.

Xia Z,Watanabe S,YamadaT, et al. (2012). Positional cloning and characterization reveal the molecular basis for soybean maturity locus E1 that regulates photoperiodic flowering. Proceedings of the National Academy Sciences, USA 109: E2155–E2164.

Yazdi-Samadi, B., Rinne, R. W., and Seif, R. D. (1977). Components of developing soybean seeds: oil, protein, sugars, starch, organic acids, and amino acids. Agronomy Journal, 69(3), 481-486.

Yuan, F.J., Zhao, H.J., Ren, X.L., Zhu, S.L., Fu, X.J. and Shu, Q.Y., (2007). Generation and characterization of two novel low phytate mutations in soybean (Glycine max L. Merr.). Theoretical and Applied Genetics, 115(7), pp.945-957.

Zeng, A., Chen, P., Shi, A., Wang, D., Zhang, B., Orazaly, M., Florez-Palacios, L., Brye, K., Song, Q. and Cregan, P., (2014). Identification of quantitative trait loci for sucrose content in soybean seed. Crop Science,54(2), pp.554-564.

Zhou, J.R., Fordyce, E.J., Raboy, V., Dickinson, D.B., Wong, M.S., Burns, R.A. and Erdman, J.W., (1992). Reduction of phytic acid in soybean products improves zinc bioavailability in rats. JOURNAL OF NUTRITION-BALTIMORE AND SPRINGFIELD THEN BETHESDA-, 122, pp.2466-2466.

Zuther E., Buchel K., Hundertmark M., Stitt M., Hincha D.K., Heyer A.G. (2004) The role of raffinose in the cold acclimation response of Arabidopsis thaliana. FEBS Lett 576: 169–173

Chapter 2

Identification of a haplotype and improved associated molecular marker for the soybean seed ultra-low RFO phenotype

Introduction

Soybean [*Glycine max* (L.) Merr.] is one of the most important economic plants in the world. Soybean consists of 20 % oil, 40 % protein, and 15 % soluble carbohydrates in the dry weight of seeds (Hsu et al., 1973). Soybean meal is a possible source of protein for monogastric animals. In 2014, 43.1 million short tons were used for animal feed and more than half of soybean meal production in the U.S. was utilized as poultry feed and a quarter of that was used to feed swine (SoyStat.com 2015). For its purpose of use, soybean seed composition used to be modified with high protein by traditional breeding method. There are three major components of the soluble carbohydrates in soybean such as sucrose, raffinose and stachyose (Openshaw and Hadley, 1978). Sucrose is the only beneficial component and is digested to use as metabolizable energy in the animal gut. However, raffinose family oligosaccharides (RFO) such as raffinose and stachyose are considered as anti-nutritional factors because those components are not able to be digested due to the lack of α-galactosidase activity in the gut of monogastric animals.

Several studies mentioned that soybean meal with low RFO was considered a beneficial animal feed with increased metabolizable energy when consumed by animals. According to one study (Coon et al., 1990), the results demonstrated that decreasing raffinose and stachyose in soybean meal had 20 % more metabolizable energy in roosters. Adding RFO components to animal feed was reported to decrease nutrient digestibility in pigs (Smiricky et al., 2002). In a chicken study, feeding soybean meal with low RFO was reported to have higher metabolizable energy than conventional soybean meal (Parsons et al., 2000). Sucrose is the only soluble carbohydrate component to be a beneficial component for animals. The studies demonstrated that soybean with decreased RFO contained an increased sucrose concentration (Parson et al., 2000;

Hitz et al., 2002; Neus et al., 2005; Obendorf et al., 2008). Soybean seeds are processed through several steps with soaking, heating and pressure to make soybean meal. However, these steps of processing for soybean meal decrease not only RFO, but also sucrose content from soybean seeds (Leske and Coon, 1999). Although soybean seeds are processed, RFO still remain. Thus, a full genetic understanding of decreasing or even removing raffinose and stachyose from soybean would help soybean have more nutritional content when consumed by animals.

The soybean plant introduction PI 200508 was shown to have a lower content of RFO and an increased content of sucrose compared to wild-type plants (Kerr and Sebastian, 2000). In addition, another study indicated that a mutant line with reduced RFO has deceased level of raffinose synthase activity and increased levels of sucrose (Hitz et al., 2002). A soybean line with a mutation in a myo-inositol phosphate-1 synthase (MIPS1) which plays an important role in the carbohydrate synthesis pathway, was found to have a reduced level of RFO as well as phytic acid (Hitz et al., 2002). But, soybean lines with the MIPS mutation had low field emergence (Meis et al., 2003). Although the mutant MIPS1 allele causes soybean lines to have a low level of RFO in the seed, it cannot be good source for soybean meal. Another mutant, *stc1* from PI 200508 resulting in a low RFO phenotype, was reported to have reduced raffinose synthase activity but normal stachyose synthase and galactinol synthase activities (Sebastian et al., 2000; Hitz et al., 2002). In addition, there were no negative associations of the *stc 1* allele with agronomic traits (Neus et al., 2005).

The genetic basis of the reduced RFO trait derived from PI 200508 was reported. They found that a novel allele contained a deletion of three base pairs within the raffinose synthase 2 coding sequence (*RS2*, Glyma06g18890, *Glycine max v1.1*), resulting in the deletion of a

tryptophan residue at position 331 (*rs2W331-*) in a highly conserved region (Dierking and Bilyeu, 2008). A mutation of this region in soybean results in the increased sucrose level and the decreased raffinose and stachyose phenotype. Using reverse genetics, another mutation in the RS2 gene was identified in soybean. This mutant allele of the RS2 gene contains a missense mutation resulting in the change of isoleucine from threonine at amino acid position 107 (*rs2T107I*) (Dierking and Bilyeu, 2009). Soybean lines with the *rs2T107I* alleles also showed reductions in raffinose and stachyose along with an increase in sucrose. However, soybean lines with the *rs2W331-* allele had a lower level of RFO than the ones with the *rs2T107I* allele (Hagely et al., 2013).

Soybean lines with significant reductions in RFO beyond the observed low RFO levels in soybean lines were discovered through traditional plant breeding, and the phenotype was designated as ultra-low RFO phenotype (Schillinger et al., 2013; Hagely et al., 2013). Soybean with ultra-low RFO content was defined as a less than 0.13 % of raffinose seed content and 1.6 % of stachyose content by dry weight of the total seed content, respectively (Schillinger et al., 2013). This ultra-low RFO phenotype contained an additional allele with the *rs2W331-* allele. A soybean genome duplication occurred so that approximately 75 % of the genes in soybean have multiple copies (Schmutz et al., 2010). Four raffinose synthase genes (RS1, RS2, RS3 and RS4) were identified in the soybean genome based on sequence homology (Dierking and Bilyeu, 2008). Of these, the additional allele was associated in a mapping population with a naturally occurring variant of the raffinose synthase 3 gene (*RS3*, Glyma05g08950) on chromosome 05 by 1536 SNP chip analysis and RS3 sequencing analysis (Hagely, 2013; Hyten et al., 2010). According to this study, RS3 was a candidate gene for the ultra-low RFO phenotype, and five

single nucleotide polymorphisms (SNP) were identified by sequencing the RS3 exon sequences of several ultra-low RFO lines. Of these, gene-based markers for the variant *rs3snp5* resulting in a silent mutation in exon 1 was developed and used for marker-assisted selection for the ultra-low RFO phenotype.

In the previous study, RS3 was an important region associated with the ultra-low RFO phenotype by the genetic mapping process (Hagely et al., 2013). As genomic sequencing information is publicly available, SNPViz software was recently developed to identify shared haplotypes from nucleotide polymorphisms among sequenced soybean lines in a user-defined region (Langewisch et al., 2014). For the present study, we had utilized this software to survey the number of haplotypes of RS3 alleles and determined a SNP position that defined a unique RS3 haplotype. To test this SNP position, we developed independent populations in our research inventory with *rs2W331-* alleles and one of three *RS3* haplotypes to associate the RFO phenotypes with a marker that distinguished the *RS3* haplotypes. Through this process, we were able to predict the development of soybean lines having the ultra-low RFO phenotype.

Materials and Methods

Plant material and segregating populations

The soybean lines Macon (PI 593258, MG III) (Nickell et al., 1996), Lincoln (PI 548362, MG III, http://www.ars-grin.gov/npgs/), Deuel (PI 662940, MG I, http://www.ars-grin.gov/npgs/), Patriot (MG III), SGUL (Bilyeu and Wiebold, 2016) and KB10-29#1639 were used in this study. The named lines are conventional soybeans used in the breeding program; SGUL and KB10-29#1639 are experimental lines developed for altered carbohydrate profiles. In addition, three genetic populations were initiated in the summer of 2012 at South Farm, University of Missouri - Columbia, MO.

Population 1 was developed from a cross between Macon and KB10-23#1639. Population 2 was developed from a cross between Lincoln and KB10-23#1639. Population 3 was developed from the cross of Deuel and SGUL. The F_1 and F_2 generations were grown in the winter of 2012/2013 in Costa Rica. In the F_2 generation, single plants were selected according to genotyping assays for *rs2W331-* and *rs3snp5* alleles. For population 3, both parental lines had the mutant allele of *rs3snp5*, so selection was only for *rs2W331-*. Single plant threshes of F_2:$_3$ seeds were made of all selected individuals. The selected individuals from all populations of F_3 were grown in the summer of 2013 at South Farm Research Center in Columbia MO. The experimental design was a randomized set of blocks each containing a single replication of a plot consisting of F_2:$_3$ plants. For analysis, three single F_3 plants from each plot were individually harvested, and two F_2:$_3$ samples of 15 seeds were used from each plot were used to measure the carbohydrates in seeds. In 2014, separate field experiments were performed at South Farm Research Center and Greenley Research Center near Novelty, MO. For the 2014 experiments,

the experiment design was a randomized complete block with three replications. Each replication consisted of ten $F_{3:4}$ seeds planted per plot by hand into 91 cm plots with 30 cm spacing, and seeds were harvested in bulk to analyze the carbohydrate components by HPLC.

DNA isolation and PCR for Sequencing

Genomic DNA was isolated from ~20 mg powdered seed tissue ground under liquid nitrogen using the DNeasy Plant Mini Kit (Qiagen Sciences Inc., Germantown, MD, U.S.) and used at 5-50 ng per PCR amplification. PCR was conducted with the following conditions: 95ºC for 3 minutes followed by 40 cycles of 95ºC for 20 s, 60ºC for 20 s, 72ºC for 20 s. PCR products were run by gel electrophoresis (1.5 % agarose gel) to identify the size of product and ensure specific amplification. PCR products were isolated with the QIAprep Spin Miniprep kit (Qiagen Sciences Inc.) and sequenced at the University of Missouri DNA Core facility. After getting the sequencing information, multiple sequencing alignment was conducted by ClusterW program to make a boxshade plot.

The RS3snp6 genotyping assay

Simpleprobes were designed using the Lightcycler Probe Design Software (Roche Applied Sciences) by input of target region of sequence with a highlighted single mutant position, and Simpleprobes were ordered from Fluoresentric, Inc., Park City, UT). The RS3snp6 genotyping assays were conducted with an asymmetric mixture of primers [0.5µM forward (5’ GTCACGCGCTTGAGAGAAAT 3’) and 0.2µM reverse (5’ TACTCCACCGACCCAAACTC

3') with 5:2 asymmetric mix of primers]; the *rs3snp6* SimpleProbe was used at a concentration of 2µM of 5' SPC-TTGCAAAAGCATTGTTATTCCCATTAAG-Phosphate 3'. The total volume of reactions was 20µl containing 5-50ng DNA template, primers, buffer (40mM Tricine-KOH (pH 8.0), 16mM KCl, 3.5mM $MgCl_2$, 3.75ug ml^{-1} BSA, 200 µM dNTPs), 5 % DMSO, SimpleProbe, and 0.2X Titanium Taq polymerase (BD Biosciences, Palo Alto, CA). Assay reactions were conducted in a 96-well plate in a Roche LightCycler 480 II using the following reaction conditions: 95ºC for 3 minutes followed by 40 cycles of 95ºC for 20 s, 60ºC for 20 s, 72ºC for 20 s and then a melting curve from 49º to 65ºC in the *rs3snp6* assay. Fluorescence was read every 0.1 ºC from 49 ºC to 65 ºC and the single nucleotide match or mismatch was discriminated by integrating the fluorescence disappearance with increasing melting temperature. The homozygous wild type allele of RS3 was detected by a peak at 62.7 ºC; the homozygous mutant allele containing snp6 of RS3 showed a peak at 56.6 ºC, and heterozygous alleles produced both peaks.

Phenotype determination by HPLC

Soluble carbohydrates in soybean were determined by high performance ion chromatography with pulsed amperometric detection (PAD) employing a Dionex ICS-5000 with Electrochemical Detector (Thermo Scientific Dionex, Waltham, MA). Around 15 seeds from all genotypes were placed in -80 °C overnight and lyophilized to dryness prior to powdering. To phenotype, soluble carbohydrates in seed were extracted by an ethanol extraction method. Briefly, 12.5 mg of sample was extracted with 1 mL of 50 % ethanol for 30 min at 70 °C with three times intermittent shaking in a 2 mL microcentrifuge tube. Then, samples were centrifuged

10 min at 16,000 x g. Approximately 700 μL of the supernatant was removed and stored at 4 °C before further experiments. A 50 μL aliquot of each sample was dried under vacuum and resuspended in 250 μL deionized water, with 10 μL injection. Samples were placed in 96 well plate and automatically applied to the column. Four soluble carbohydrates (galactinol, sucrose, raffinose, and stachyose) were separated on a Dionex Carbo Pac PA 10 analytical column (250 mm x 4 mm, 10 μm) connected to a Carbo Pac PA 10 guard column (50 nm x 4 nm). The mobile phase was 90 mM NaOH (blanketed with helium) with flow rate of 1.5 ml min-1. A gold electrode was used in the electrochemical cell of the detector, and the settings were (Time in seconds/Volts): 0/0.1; 0.2/0.1; 0.4/0.1; 0.41/-2.0; 0.42/-2.0; 0.43/0.6; 0.44/-0.1; 0.5/-0.1. Run time was a total of 48 minutes, with the first 18 minutes for sample separation followed by a 15 minutes washing step with 200 mM NaOH, and a 15 minutes re-equilibration step with 90 mM NaOH. Peak areas were integrated for galactinol, sucrose, raffinose, and stachyose. Carbohydrates were quantified based on standard curves generated for each carbohydrate. The content of galactinol, sucrose, raffinose, and stachyose is reported as the percent of dry seed weight, which can be converted to $g \cdot kg^{-1}$ by multiplying the percent of dry seed result by ten.

Results

Previous research led to the discovery of a SNP resulting in a silent mutation in the first exon of the RS3 gene on chromosome 05 that was associated with the ultra-low RFO phenotype when the *rs2W331*- mutation was present (Schillinger et al., 2013). A molecular marker assay was developed for the SNP (Hagely, 2013), and because it was the fifth SNP present in a small survey of soybean RS3 gene sequences, it was named RS3snp5. While not present in the plant introduction line PI 200508 or the reference sequence of Williams 82, based on the molecular marker assay, the RS3snp5 variant allele was present in several U.S. soybean cultivars and a subset of the North American soybean ancestor accessions (Figure 2.1). With the publication of 31 additional soybean genome sequences, the availability of the soybean Nested Associated Mapping (NAM) parent sequences, and the use of the SNPViz haplotype visualization tool, it was determined that there were actually two different haplotypes of RS3 that each contained RS3snp5 but were distinct from each other and the reference Williams 82 sequence (Figure 2.1) (Lam et al., 2010; Langewisch et al., 2014). The Williams 82 haplotype of the RS3 gene region is referred to here as RS3 haplotype A. The RS3snp5 G/T SNP on chromosome 05 was at position 8,810,126 in *Glycine max v1.1*. A different SNP (C/T in intron 1) was identified on chromosome 05 at position 8,809,816 that distinguished the two RS3snp5-containing haplotypes from each other and from the reference sequence (Figure 2.1). Lines with only the RS3snp5 SNP are referred to here as RS3 haplotype B, while lines with both the RS3snp5 and RS3snp6 variants are referred to here as RS3 haplotype C (Figure 2.1).

The Williams 82 RS3 sequence is coded here as haplotype A, indicating the gene has wild-type reference sequence for the RS3 alleles and is presumably fully functional. To identify

soybean accessions in our research program that contained different alleles of the RS3 gene, DNA extraction and sequencing of part of the RS3 gene was conducted for several soybean lines which contained RS3snp5 variant alleles (Figure 2.2. a). The genomic sequence data used for SNPViz contains the RS3 gene in the negative sense orientation, but our analyses of the gene sequence is in the positive sense orientation. According to the alignment of the sequencing results, the RS3snp6 position for Williams 82 as a reference sequence has a "G" nucleotide corresponding to genomic Williams 82 chromosome 05 position 8,809,816, which was consistent with Lincoln and Deuel, thus both RS3 haplotype B; Soybean cultivars 534545, Macon, Patriot, SGUL and IA3023 have an adenine instead of guanine and are thus RS3 haplotype C (Figure 2.1; Figure 2.2.a).

To understand the phenotypic impact of the different RS3 haplotypes, we carried out a genotype/phenotype association analysis with three different segregating populations (Population 1, 2 and 3). For populations 1 and 2, KB12-23 #1639 (*rs2W331-*/RS3, haplotype A) had a low RFO phenotype and was the donor of *rs2W331-* alleles with Macon (RS2/RS3, haplotype C) and Lincoln (RS2/RS3, haplotype B). For these populations, in the F_2 generation, individual plants having homozygous *rs2W331-* and *rs3snp5* alleles were identified by molecular marker genotyping assays. Seeds (F_4) from the identified F_2 plants were produced in a Missouri field environment in 2013, and harvested seeds were analyzed by HPLC to identify the carbohydrate profile including galactinol, sucrose, raffinose and stachyose. The results from population 1 (fixed for *rs2W331-* and segregating for RS3 haplotypes A and C) showed individual lines with distinctly different phenotypes of either low RFO or ultra-low RFO over two years (Figure 2.4; Figure 2.5). For population 2 (fixed for *rs2W331-* and segregating for RS3 haplotypes A and B),

both individuals evaluated showed the low RFO phenotype with no ultra-low RFO phenotypes over years (Figure 2.4; Figure 2.5). According to these results, the cultivar Macon with the RS3 haplotype C is a possible donor for the ultra-low RFO trait, while the ancestral line Lincoln, with the RS3 haplotype B does not contribute to the ultra-low RFO trait. For further confirmation, population 3 with SGUL as the donor of *rs2W331-* and RS3 haplotype C alleles and Deuel with Deuel as a donor of RS3 haplotype B alleles was identified as an additional test of the impact of RS3 haplotype B and C. As in populations 1 and 2, in the F_2 generation, individual plants having homozygous *rs2W331-* and *rs3snp5* alleles were identified by molecular marker genotyping assays. Seeds (F_4) from the identified F_2 plants were produced in a Missouri field environment in 2013, and harvested seeds were analyzed by HPLC to identify the carbohydrate profile including galactinol, sucrose, raffinose and stachyose. Lines from population 3 showed either the low RFO or ultra-low RFO phenotype result. Thus, we concluded that ultra-low RFO phenotype is not perfectly associated with *rs3snp5*.

A molecular marker assay was developed to detect the variant SNP on chromosome 05 at position 8,809,816, referred to as RS3snp6, which defines the RS3 haplotype C. A simpleProbe-based assay was designed to distinguish RS3 haplotypes A and B containing the reference Williams 82 sequence on chromosome 05 at position 8,809,816 from the variant base present in RS3 haplotype C (figure 2.3.b). Although it is a SNP, this high throughput assay was able to efficiently distinguish the variant allele from the reference allele by analyzing the melting curve temperature after the PCR was completed. The peak of the variant allele was detected at 56.6 °C, whereas the peak of the reference allele was at 62.7 °C. Heterozygous lines were also determined to have two peaks of melting temperature (Figure 2.3.c).

Using the RS3snp6 molecular marker assay, the lines selected from populations 1, 2, and 3 were genotyped to assign their RS3 haplotype based on the RS3snp5 and RS3snp6 status (Table 2.2). The subset of selected lines from the three populations with parental lines were grown in the summer of 2014 in two Missouri locations in a replicated trial, and carbohydrate profiles of seeds were determined with HPLC to analyze galactinol, sucrose, raffinose and stachyose (Figure 2.5). Williams 82 (RS2/RS3 haplotype A) was used as the control carbohydrate profile. The carbohydrate profiles of lines containing reference RS2 alleles, Macon, Lincoln, and Deuel showed a result similar to the Williams 82 phenotype (Figure 2.5; Table 2.1) despite the differences in RS3 haplotypes. KB10-23#1639 had a low RFO phenotype which matched with the *rs2W331-/RS3* haplotype A genotype (Figure 2.5; Table 2.1). Since SGUL has the *rs2W331-* alleles and the *rs3snp5* and *rs3snp6* alleles defining RS3 haplotype C, the phenotype was a carbohydrate profile containing a minimal amount of RFO. Population 1 showed two distinct carbohydrate profiles: ultra-low RFO for all RS3 haplotype C lines and low RFO for all RS3 haplotype A lines. For population 2, both lines were RS3 haplotype B and had low RFO phenotypes. For population 3, lines having RS3 haplotype C had the ultra-low RFO phenotype, whereas individuals with RS3 haplotype B showed low RFO phenotypes (Figure 2.5). The 2013 and 2014 results showed very similar carbohydrate profiles for each line (Figure 2.4; Figure 2.5). Additionally, the RS2 and RS3 genotype could be perfectly matched with the carbohydrate phenotype when utilizing the *rs3snp6* marker, but not with the *rs3snp5* marker.

The results from the selected lines that are fixed for the *rs2W331-* alleles in populations 1, 2, and 3 demonstrate the effect of RS3 haplotype C in significantly reducing seed RFO when compared to lines that are fixed for the *rs2W331-* alleles and contain either the RS3 haplotype A

or B (figure 2.6). When combining all of the lines in populations 1, 2, and 3 by RS3 haplotype (all contain *rs2W331-* alleles), the *RS3* haplotype C lines in the three populations have a statistically similar stachyose content compared to parent 5 (ultra-low RFO) and significantly lower stachyose content than RS3 haplotypes A or B. Lines with RS3 haplotypes A and B were not statistically different for stachyose content compared to the low RFO parent 3. Lines from the three populations combined by RS3 haplotype had no statistically different raffinose content compared with parent 3 (low RFO) and parent 5 (ultra-low RFO) (Figure 2.6). Parental lines containing haplotype B and C had significantly higher levels of raffinose and stachyose because they have functional RS2 alleles (Figure 2.6).

Discussion

Many researchers were interested in the development of soybean lines with a reduction in the amount of anti-nutritional factors for human and animals. RFO is one of the anti-nutritional factors because they are unable to be digested in the gut of monogastric animals. The soybean plant introduction PI 200508 was shown to have a lower content of RFO and an increased content of sucrose compared to wild-type plants (Kerr and Sebastian 2000). Genetic analysis of PI 200508 suggested that the low RFO phenotype was controlled by RS2 gene containing a three base pair deletion which caused a tryptophan deletion on 331 position of the amino acid sequence (*rs2W331-*), which is a recessive allele (Dierking and Bilyeu, 2008). In addition, the first examples of soybean lines with significant reductions in RFO beyond PI 200508 in soybean lines, called the ultra-low RFO phenotype were discovered and also had *rs2W331-* alleles present (Schillinger et al., 2013; Hagely et al., 2013).

Previous work was conducted with genetic mapping studies with the use of a SNP chip and candidate gene work to identify a new allele of the RS3 gene associated with ultra-low RFO phenotype when in conjunction with the RS2 allele (Hagely, 2013; Hyten et al., 2010). According to this research, they had discovered a SNP position resulting in a silent mutation in the first exon of the RS3 gene on chromosome 05 at position 8,810,126 in *Glycine max v1.1* (Schillinger et al., 2013). With this information, a molecular marker assay was developed (RS3snp5) (Hagely, 2013).

As genomic sequencing information is publicly available, SNPViz software was recently developed when comparing nucleotide polymorphisms, SNP positions and allelic variations across many soybean samples (Langewisch et al., 2014). We had surveyed the haplotype of the

RS3 region with 31 additional soybean genome sequences and soybean Nested Associated Mapping (NAM) parent sequences and localized the RS3snp5 at position 8,810,126 on chromosome 05. With the use of SNPViz haplotype visualization software, it was determined that there were actually two different haplotypes of RS3 that each contained RS3snp5 but were distinct from each other and the reference Williams 82 sequence. Three different haplotypes of RS3 gene region were identified through haplotype visualization with multiple sequences of soybean lines.

As a result of SNPViz software, the base pair differences between wild type and *rs3snp6* alleles from a "G" to an "A" was found in intron 1 of RS3 gene which is 310 base pair away from *rs3snp5* allele in exon 1 of RS3 gene. This SNP position was verified by sequence analysis for several soybean lines corresponding to position 8,809,816 on chromosome 05 by multiple sequencing alignment. A high throughput genotyping assay utilizing SimpleProbe chemistry and real-time PCR can distinguish a single base pair difference between wild type and mutant alleles by determination of melting temperature.

Although no causative mutant polymorphisms were discovered that explain the molecular basis for the haplotype C variant RS3 alleles to associate with the ultra-low RFO phenotype, this work improved the selection potential for the phenotype with the development and validation of the *rs3snp6* molecular marker assay. We determined that the RS3 haplotype C alleles are present in modern breeding lines that can be utilized in combination with *rs2W331-* alleles to produce the ultra-low RFO phenotype in soybean seeds. Through this work, the new marker assay can be utilized in traditional plant breeding programs to determine the use of parental lines for crossing

combinations with RS3 haplotype C and make genotypic selections for the ultra-low RFO phenotype.

Tables

Table 2.1. Nomenclature, RS2 and RS3 Allelic Status (*rs3snp5* and *rs3snp6*)

Name	Genotype				Note
	RS2 allele	RS3snp5	RS3snp6	RS3 haplotype	
Macon	RS2[1]	*rs3snp5*[2]	*rs3snp6*[3]	C[4]	Inbred (Parent 1)
Lincoln	RS2	*rs3snp5*	RS3	B	Inbred (Parent 2)
KB10-23#1639	*rs2W331-*	RS3	RS3	A	Low RFO[5] (Parent 3)
Deuel	RS2	*rs3snp5*	RS3	B	Inbred (Parent 4)
SGUL	*rsW331-*	*rs3snp5*	*rs3snp6*	C	Ultra-low RFO (Parent 5)
KB12-20 (Population1)	*rs2W331-*			A/C	Parent 1 x Parent 3
KB12-21 (Population2)	*rs2W331-*			A/B	Parent 2 x Parent 3
KB12-31 (Population3)	*rs2W331-*			B/C	Parent 4 x Parent 5

[1]RS2 indicates functional raffinose synthase 2 gene which is similar to Williams 82 reference genome for Glyma06g18890. *rs2W331-* indicate the mutant alleles (Dierking and Bilyeu 2008). [2]RS3snp5 indicates raffinose synthase 3 gene which is similar to reference genome for Glyma05g08950. RS3snp5 is the mutant allele of RS3 (Hagely, 2013). [3]RS3snp6 indicates a new molecular marker for this study. [4]The RS3 haplotype where A is wild-type RS3 for both RS3snp5 and RS3snp6 like Williams 82; B is RS3snp5 RS3snp6 with variant *rs3snp5* and wild-type RS3snp6, and C is variant for both *rs3snp5* and *r3snp6*. [5]A reduction of raffinose and stachyose like PI200508 (Hitz et al., 2002).

Table 2.2. Nomenclature, RS2 and RS3 Allelic Status (*rs3snp5* and *rs3snp6*)

Populations	Genotype				
	RS2 allele	RS3snp5	RS3snp6	Haplotype	Note
KB12-20 (Population 1)					
KB12-20 #703	*rs2W331-*[1]	*rs3snp5*[2]	*rs3snp6*[3]	C[4]	Ultra-low RFO[5]
KB12-20 #669	*rs2W331-*	*rs3snp5*	*rs3snp6*	C	Ultra-low RFO
KB12-20 #670	*rs2W331-*	RS3	RS3	A	Low RFO[6]
KB12-20 #687	*rs2W331-*	RS3	RS3	A	Low RFO
KB12-20 #620	*rs2W331-*	RS3	RS3	A	Low RFO
KB12-20#770	*rs2W331-*	RS3	RS3	A	Low RFO
KB12-21 (Population 2)					
KB12-21 #708	*rs2W331-*	*rs3snp5*	RS3	B	Low RFO
KB12-21 #767	*rs2W331-*	*rs3snp5*	RS3	B	Low RFO
KB12-31 (Population 3)					
KB12-31 #974	*rs2W331-*	*rs3snp5*	RS3	B	Low RFO
KB12-31 #1039	*rs2W331-*	*rs3snp5*	RS3	B	Low RFO
KB12-31#1042	*rs2W331-*	*rs3snp5*	RS3	B	Low RFO
KB12-31 #999	*rs2W331-*	*rs3snp5*	*rs3snp6*	C	Ultra-low RFO
KB12-31 #1004	*rs2W331-*	*rs3snp5*	*rs3snp6*	C	Ultra-low RFO
KB12-31 #1027	*rs2W331-*	*rs3snp5*	*rs3snp6*	C	Ultra-low RFO
KB12-31 #1008	*rs2W331-*	*rs3snp5*	*rs3snp6*	C	Ultra-low RFO

[1]RS2 indicates functional raffinose synthase 2 gene which is similar to Williams 82 reference genome for Glyma06g18890. *rs2W331-* indicate the mutant alleles (Dierking and Bilyeu 2008). [2]*rs3snp5* indicates raffinose synthase 3 gene which is similar to reference genome for Glyma05g08950. *rs3snp5* is the mutant allele of RS3 (Hagely, 2013). [3]*rs3snp6* indicates a new molecular marker for this study. [4]The RS3 haplotype where A is wild-type RS3 for both

RS3snp5 and RS3snp6 like Williams 82; B is RS3snp5 RS3snp6 with variant *rs3snp5* and wild-type RS3snp6, and C is variant for both *rs3snp5* and *r3snp6*. [5]Minimal amount of RFO (Hagely et al., 2013). [6]A reduction of raffinose and stachyose like PI200508 (Hitz et al., 2002).

Figures

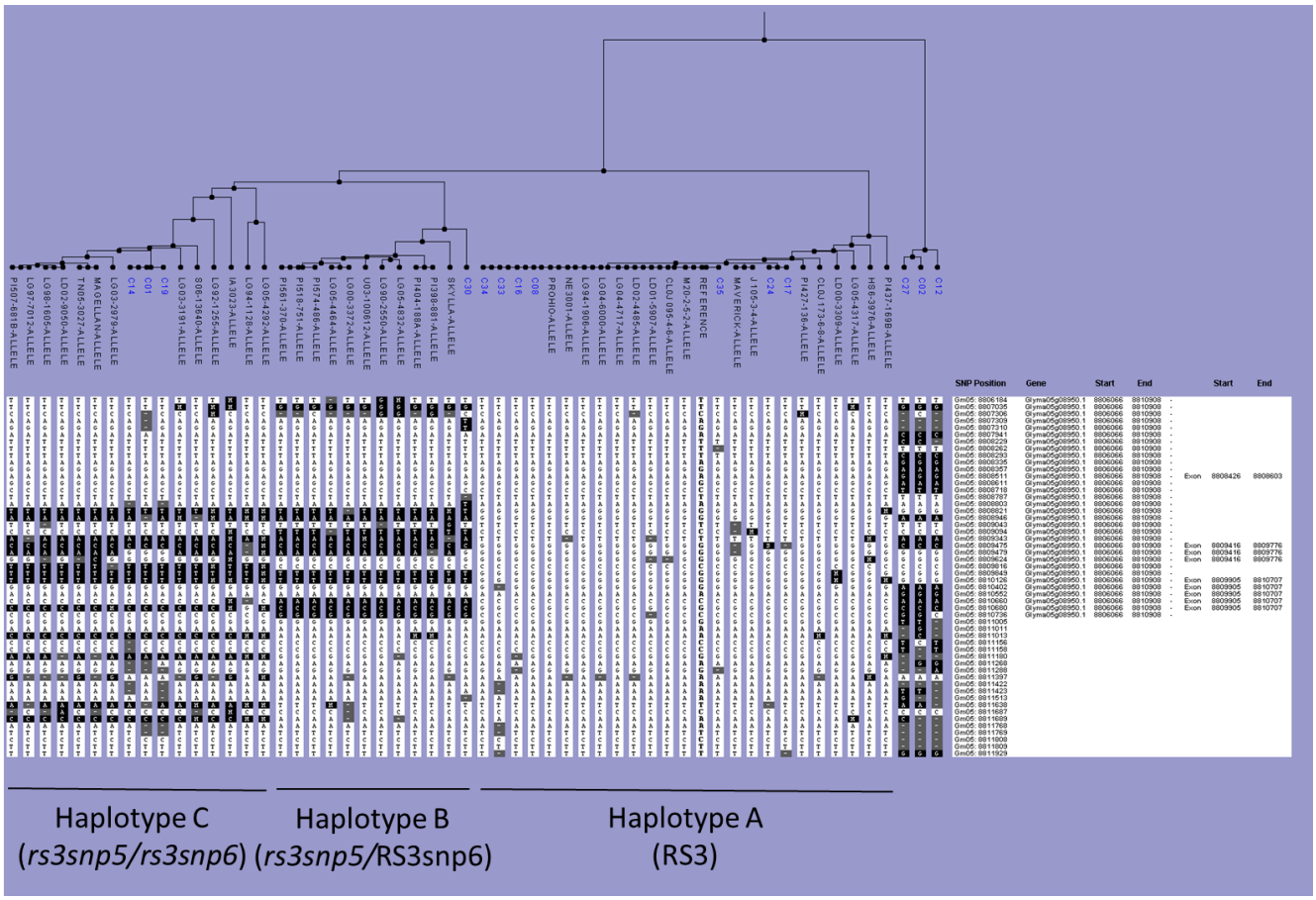


Figure 2.1. SNPViz haplotype analysis of the RS3 gene region. A 5.8Kb region on chromosome 5 is presented including the RS3 allele region (Glyma05g08950) with phylogenic tree across multiple sequenced soybean lines. The reference is the sequence of Williams 82. Blue font represents sequencing of 14 cultivated soybean lines (Lam et al., 2010) and black font means nested association mapping (NAM) parents. Only positions that have at least one polymorphism within the dataset are indicated. When an accession has the identical nucleotide with the reference sequence, the position is represented with a white background; the black background indicates a polymorphic single nucleotide compared to the reference, and missing data is represented as gray highlight. The actual nucleotide at each position is indicated. The RS3 haplotype where A is wild-type RS3 for both RS3snp5 and RS3snp6 like Williams 82; B is

RS3snp5 RS3snp6 with variant *r3snp5* and wild-type RS3snp6, and C is variant for both *rs3snp5* and *r3snp6*.

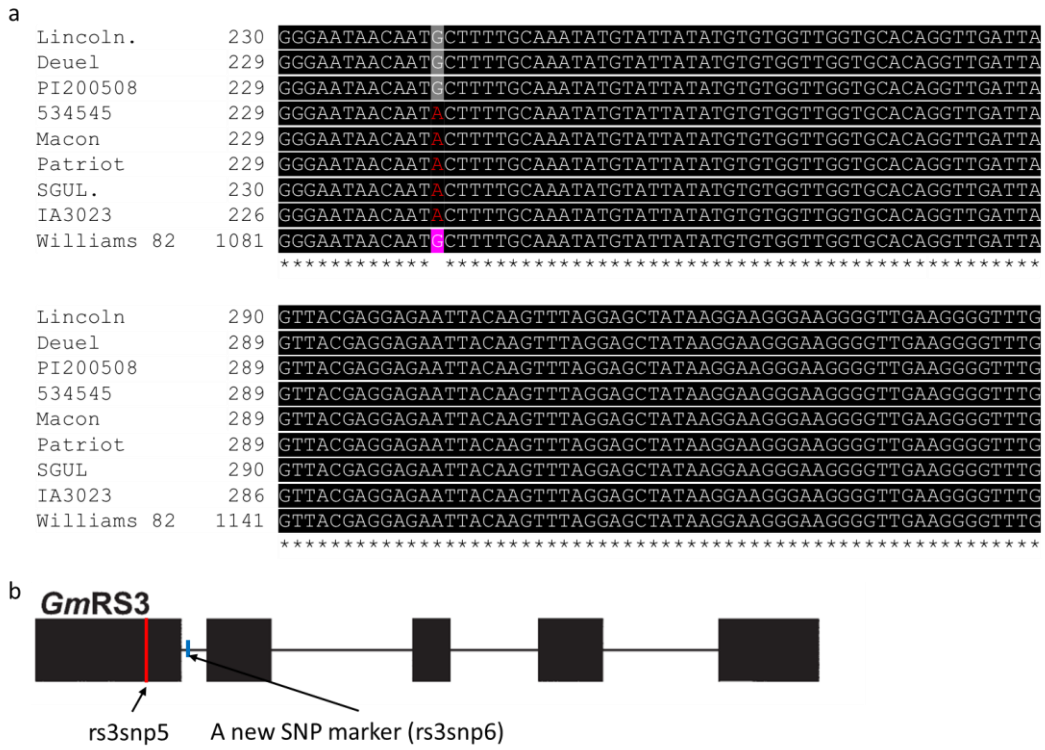


Figure 2.2. Multiple sequencing alignments and Markers position in RS3 gene scheme. a. Boxshade of a ClustalW sequence alignment of part of RS3 gene with soybean lines which confirm the new SNP position; Lincoln (*rs3snp5*, Haplotype B), Deuel (*rs3snp5*, Haplotype B), PI 200508 (*rs3snp5*, Haplotype D), Macon *(rs3snp5/rs3snp6*, Haplotype C), Patriot (*rs3snp5/rs3snp6*, Haplotype C), SGUL (*rs3snp5/rs3snp6*, Haplotype C), IA3023 (*rs3snp5/rs3snp6*, Haplotype C) and reference genome Williams 82 (Haplotype A). b. Schematic genomic structure of RS gene; A red lines indicates *rs3snp5* position and a blue line is the position of a new marker (*rs3snp6*).

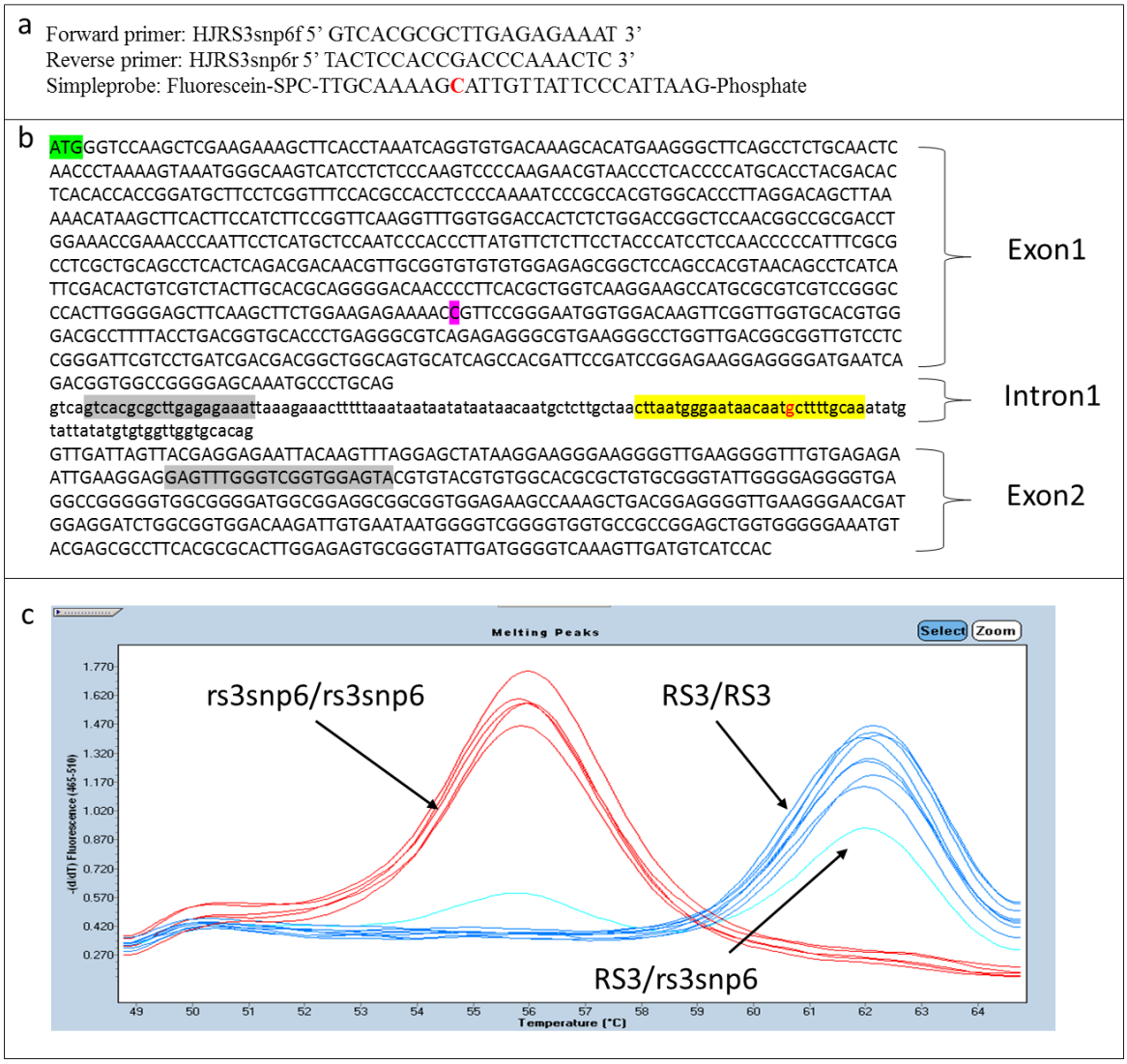


Figure 2.3. Molecular marker assay for *rs3snp6* allele. a. Primers and SimpleProbe sequencing information; Red letter (C) represents the position of single nucleotide polymorphism. b. Part of genome sequencing of raffinose synthase gene 3 (RS3) including exon 1, intron 1 and exon 2; A green box represents start codon of RS3 gene; A pink box is the position of *rs3snp5*; Gray boxes mean primers; Yellow box represents SimpleProbe and Red letter in this box is the position of SNPs for *rs3snp6* assay. c. Peaks indicate melting temperatures of wild type RS3 and mutant *rs3snp6* allele.

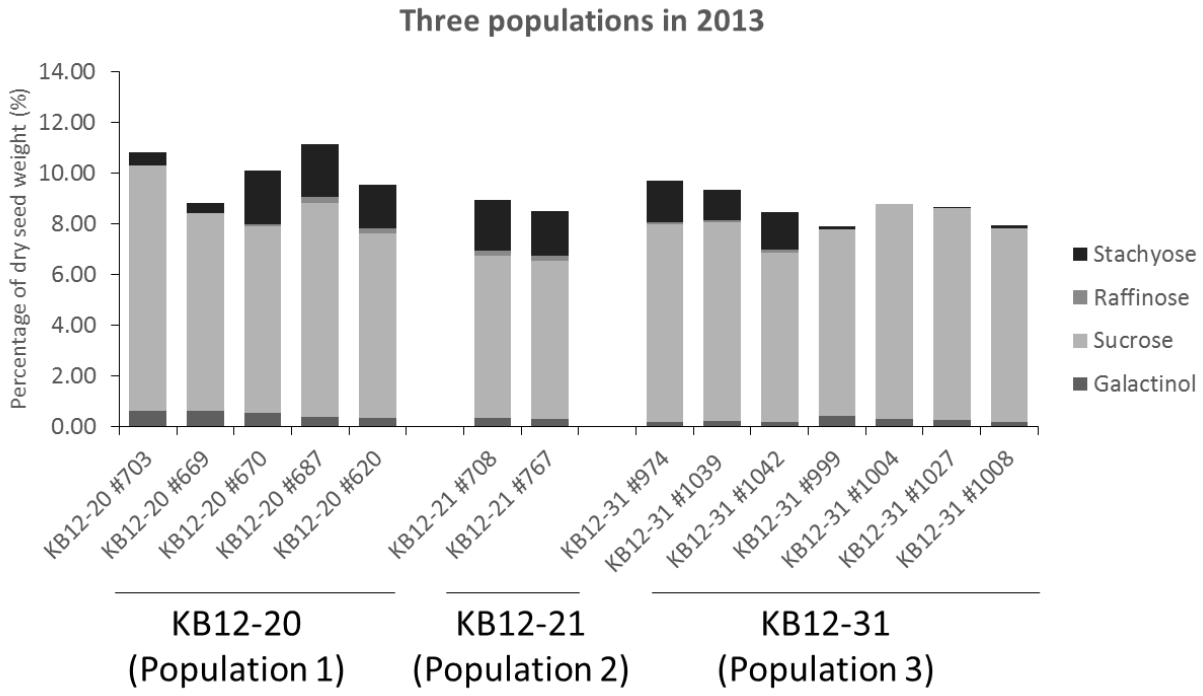


Figure 2.4. Comparison of overall carbohydrate profiles for three different populations in 2013. The average of each carbohydrate is represented by different colors in the stacked histogram. The amount of each carbohydrate component was measured as a percentage of carbohydrate in soybean seed. Three populations were tested by HPLC and two of those populations (KB12-20 and KB12-31) were showed segregating phenotype such as low RFO or ultra-low RFO phenotype.

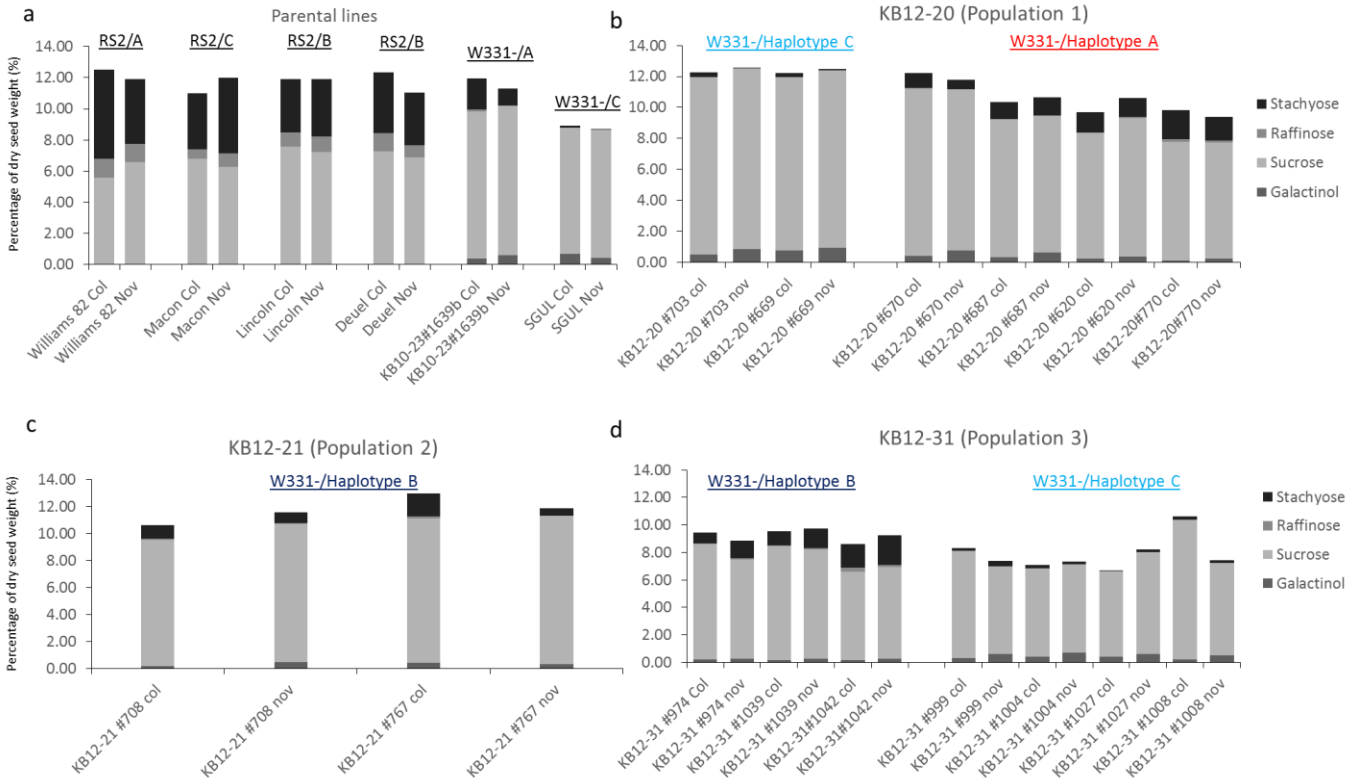


Figure 2.5. Comparison of overall carbohydrate profiles for three different populations in 2014. The average of each carbohydrate is represented by different colors in the stacked histogram. The amount of each carbohydrate component was measured as a percentage of carbohydrate in soybean seed. Three populations were tested by HPLC and two of those populations (KB12-20 and KB12-31) were showed segregating phenotype such as low RFO or ultra-low RFO phenotype, but RS3 haplotype of low RFO phenotype was different.

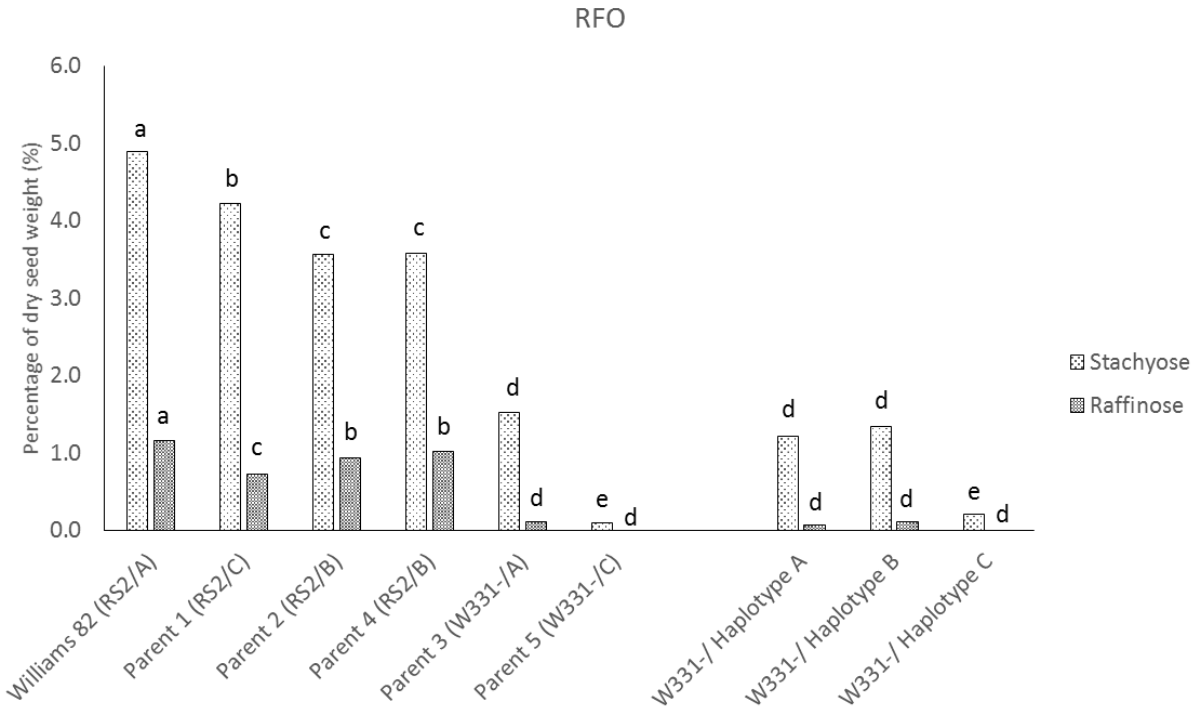


Figure 2.6. Content of raffinose and stachyose of soybean lines including parental lines and three RS3 haplotypes (A, B and C) averaged across two locations with three replications from three populations. Letter in the brackets represented three RS3 haplotype (A, B and C) with RS2 allele status in parental lines. The same letter above each bar within raffinose and stachyose indicate no significant difference at $p \leq 0.05$.

References

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.

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

Dierking, E. C., & Bilyeu, K. D. (2009). New sources of soybean seed meal and oil composition traits identified through TILLING. BMC plant biology, 9(1), 89.

Dierking, E. C., & Bilyeu, K. D. (2008). Association of a soybean raffinose synthase gene with low raffinose and stachyose seed phenotype. The Plant Genome, 1(2), 135-145

Hagely, K. B., Palmquist, D., & Bilyeu, K. D. (2013). Classification of distinct seed carbohydrate profiles in soybean. Journal of agricultural and food chemistry, 61(5), 1105-1111.

Hagely, K. B. (2013). Improving soybean seed composition through molecular breeding for Raffinose family oligosaccharides, lectin, and trypsin inhibitors (Doctoral dissertation, University of Missouri).

Hitz, W.D., T.J. Carlson, P.S. Kerr, and S.A. Sebastian. (2002). Biochemical and molecular characterization of a mutation that confers a decreased raffinosaccharide and phytic acid phenotype on soybean seeds. Plant Physiology 128:650-660.

Hsu, S.H., Hadley, H.H. & Hymowitz, T. (1973). Changes in carbohydrate contents of germinating soybean seeds. Crop Science 13:407-410

Hyten, D. L., Choi, I. Y., Song, Q., Specht, J. E., Carter, T. E., Shoemaker, R. C., ... & Cregan, P. B. (2010). A high density integrated genetic linkage map of soybean and the development of a 1536 universal soy linkage panel for quantitative trait locus mapping. Crop science, 50(3), 960-968.

Kerr, P. S. and Sebastian S. A. (2000). Soybean products with improved carbohydrate composition and soybean plants. United States, E. I. du Pont de Nemours and Company. 6,147,193.

Parsons, C.M., Y. Zhang, and M. Araba. (2000). Nutritional evaluation of soybean meals varying in oligosaccharide content. Poultry Science 79:1127-1131.

Lam, H. M., Xu, X., Liu, X., Chen, W., Yang, G., Wong, F. L., ... & Li, J. (2010). Resequencing of 31 wild and cultivated soybean genomes identifies patterns of genetic diversity and selection. Nature genetics, 42(12), 1053-1059.

Langewisch, T., Zhang, H., Vincent, R., Joshi, T., Xu, D. and Bilyeu, K., (2014). Major soybean maturity gene haplotypes revealed by SNPViz analysis of 72 sequenced soybean genomes. PloS one, 9(4), p.e94150.

Leske, K. and C. Coon (1999). Nutrient content and protein and energy digestibilities of ethanol extracted, low alpha-galactoside soybean meal as compared to intact soybean meal. Poult Sci. 78(8): 1177-1183.

Meis, S. J., W. R. Fehr and S. R. Schnebly (2003). Seed source effect on field emergence of soybean lines with reduced phytate and raffinose saccharides. Crop Sci. 43(4): 1336-1339.

Neus, J.D., W.R. Fehr, and S.R. Schnebly. (2005). Agronomic and seed characteristics of soybean with reduced raffinose and stachyose. Crop Science 45:589-592.

Nickell, C. D., Thomas, D. J., Cary, T. R., Heavner, D., (1996) Registration of 'Macon' soybean. Crop Science 36(5): 1410

Openshaw, S.J., and H.H. Hadley. (1978). Maternal effects on sugar content in soybean seeds. Crop Science 18:581-584.

Obendorf, R. L., A. D. Zimmerman, P. A. Ortiz, A. G. Taylor and S. R. Schnebly (2008). Imbibitional chilling sensitivity and soluble carbohydrate composition of low raffinose, low stachyose soybean seed. Crop Sci. 48(6): 2396-2403.

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

SoyStats (2015) Available at http://www.soystats.com.

Smiricky, M.R., C.M. Grieshop, D.M. Albin, J.E. Wubben, V.M. Gabert, and G.C. Fahey, Jr. (2002). The influence of soy oligosaccharides on apparent and true ileal amino acid digestibilities and fecal consistency in growing pigs. Journal of Animal Science. 80:2433-2441.

Schillinger JA, Dierking EC, Bilyeu KD (2013) Soybeans having high germination rates and ultra-low raffinose and stachyose content. In USP Office, ed, USPTO, Vol 8,471,107. Schillinger Genetics, Inc. The United States of America as Represented by the Secretary of Agriculture, United States of America

Schmutz J., Cannon S.B., Schlueter J., Ma J., Mitros T., Nelson W., Hyten D.L., Song Q., Thelen J.J., Cheng J. (2010) Genome sequence of the palaeopolyploid soybean. Nature 463:178-183.

Chapter 3

Environmental Stability of Carbohydrate Profiles in Different Soybean Genotypes

Introduction

Soybean is an important economic crop, providing edible oil (about 20 %), protein (about 40 %) and soluble carbohydrate (about 15 %) for both human and livestock consumption. The oil, protein and carbohydrate composition can be modified to enhance functionality. The demand for soybean products has also risen with increases in the world's population and in the need for animal feeds. In the U.S., one of the main products for animal rations is protein-containing feed. Soybean is an excellent source of protein for animal feeds. The United States produced 43.1 million short tons of soybean meal in 2014. Poultry consumes the majority of soymeal, using 49 % of US soymeal, followed by swine and beef, which consume 26 % and 11 %, respectively (http://www.soystats.com).

Soluble carbohydrates in soybean consist of sucrose, raffinose and stachyose. However, raffinose and stachyose cannot be digested by humans and monogastric animals, whereas sucrose is a nutritionally useful component in human or animal diets. Raffinose and stachyose are referred to as Raffinose Family of Oligosaccharides (RFO). RFO are derived from sucrose by the subsequent addition of galactose and chain elongation during the RFO biosynthesis pathway; the reactions are catalyzed by raffinose synthase (RS) and stachyose synthase (Peterbauer and Richter 2001). RFO are considered anti-nutritional carbohydrates. RFO cannot be digested by humans and monogastric animals due to the lack of α-galactosidase which can break the structure of RFO. However, RFO are microbially fermented in the lower gut resulting in flatulence and reduced feed efficiency (Gitzelmann and Auricchio 1965; Ruttloff et al., 1967; Price et al., 1988; Sebastian et al., 2000). Consumption of soybean seed products with low levels of RFO reduced flatulence in humans (Suarez et al., 1999) and increased metabolizable energy

efficiency in chickens (Parsons et al., 2000), pigs (Smiricky et al., 2002) and dog (Zuo et al., 1996). The removal of RFO from soybean meal resulted in a net increase in metabolizable energy (Suarez et al., 1999; Parsons et al., 2000; Coon et al., 1990).

The soybean plant introduction PI 200508 was shown to have a lower content of RFO and an increased content of sucrose compared to wild-type plants (Kerr and Sebastian 2000). In addition, another study indicated that a mutant line with reduced RFO has increased levels of sucrose (Hitz et al., 2002). The genetic basis of the reduced RFO trait derived from PI 200508 was reported. A novel allele contained a deletion of three base pairs within the raffinose synthase 2 coding sequence (*RS2*, Glyma06g18890, *Glycine max v1.1*), resulting in the deletion of a tryptophan residue at position 331 (*rs2W331-*) in a highly conserved region (Dierking and Bilyeu 2008). A mutation of this region in soybean was associated with the increased sucrose level and the decreased raffinose and stachyose phenotype. Using reverse genetics, another mutation in the *RS2* gene was identified in soybean. This mutant allele of the *RS2* gene contains a missense mutation resulting in the change of isoleucine from threonine at amino acid position 107 (*rs2T107I*) (Dierking and Bilyeu 2009). Soybean lines with the *rs2T107I* alleles also showed reductions in raffinose and stachyose along with an increase in sucrose. However, soybean lines with *rs2W331-* alleles had a lower level of RFO than the ones with the *rs2T107I* alleles (Hagely et al., 2013). Also, soybean lines with significant reductions in RFO beyond the observed raffinose and stachyose levels in soybean lines were discovered through traditional plant breeding, and the phenotype was designated as the ultra-low RFO phenotype (Schillinger et al., 2013). Soybean lines containing this ultra-low RFO phenotype possessed the *rs2W331-* alleles plus an additional genetic factor. The additional factor was associated with a mutation in

raffinose synthase 3 (*RS3*, Glyma05g08950) (Hagely 2013). Gene-based markers were developed and used for marker-assisted selection for distinct carbohydrate profiles.

Soybean lines that had distinct carbohydrate profile phenotypes with different combinations of alleles of the *RS2* and *RS3* genes were previously tested in one location (Hagely et al., 2013). This study indicated that the carbohydrate profile of soybean was somewhat variable between genotypes. However, it can be distinguishable from the carbohydrate profile of seeds from each genotypic classes (Hagely et al., 2013). Another group had reported the carbohydrate profile of seven soybean accessions grown at three very different geographic locations (Kumar et al., 2010). The results demonstrated that the sucrose content increased with cooler growing conditions. In addition, another study was conducted with soybean *RS2* and *RS3* genotypes tested in five different planting dates. They demonstrated that soybean lines in late planting dates have more content of sucrose than ones in early planting dates. They suggested that higher sucrose contents were able to be produced when there were cooler temperatures during the pod filling period (Bilyeu and Wielbold, 2016). In this study, we more thoroughly evaluated soybean lines with additional combinations of the RS2 and RS3 alleles in two locations over two years to broaden the understanding of environmental impacts on seed carbohydrate composition.

Materials and Methods

Soybean genotype.

To analyze soluble carbohydrate profiles, fourteen soybean genotypes were tested in this study, including six different reduced RFO genotypic classes based on the combination of RS2 and RS3 mutant alleles. The mutant alleles in RS2 included *rs2W331-* and *rs2T107I,* and the variant allele of RS3 was identified by a SNP position (*rs3snp6)*. Williams 82 with normal RFO content was used as a check. Modified RFO profile lines were organized by RS2 and RS3 alleles into different classes as following; RS2/RS3, RS2/*rs3snp6*, *rs2T107I*/RS3, *rs2T107I*/*rs3snp6*, *rs2W331-*/RS3, and *rs2W331-*/*rs3snp6,* and a tofu commercial line (RS2/*rs3snp6*).

Growth condition.

Ten seeds of each of the soybean lines were planted by hand into 91 cm plots with 30 cm spacing. The study set was planted in South Farm Research Center near Columbia, MO (Central Missouri) and Lee Greenley Jr. Memorial Research Center near Novelty, MO (Northern Missouri) over two years. The experimental design was a randomized complete block design with three replications.

Field emergence.

Field emergence tests were conducted with hand planting 50 seeds of each soybean genotype with three replications on June 18, 2014 at South Farm Research Center in Columbia, MO. The experimental design was a randomized complete block design with three replications.

Emergence percentage was determined by counting the number of emerged plants in each plot on 11 days after plating.

Phenotype determination by HPLC

Soluble carbohydrates in soybean were determined by high performance ion chromatography with pulsed amperometric detection (PAD) employing a Dionex ICS-5000 with Electrochemical Detector (Thermo Scientific Dionex, Waltham, MA). Around 15 seeds from all genotypes were placed in -80 °C overnight and lyophilized to dryness prior to powdering. To phenotype, soluble carbohydrates in seed were extracted by an ethanol extraction method. Briefly, 12.5 mg of sample was extracted with 1 mL of 50 % ethanol for 30 min at 70 °C with three times intermittent shaking in a 2 mL microcentrifuge tube. Then, samples were centrifuged 10 min at 16,000 x g. Approximately 700 µL of the supernatant was removed and stored at 4 °C before further experiments. A 50 µL aliquot of each sample was dried under vacuum and resuspended in 250 µL deionized water, with 10 µL injection. Samples were placed in 96 well plate and automatically applied to the column. Four soluble carbohydrates (galactinol, sucrose, raffinose, and stachyose) were separated on a Dionex Carbo Pac PA 10 analytical column (250 mm x 4 mm, 10 µm) connected to a Carbo Pac PA 10 guard column (50 nm x 4 nm). The mobile phase was 90 mM NaOH (blanketed with helium) with flow rate of 1.5 ml min-1. A gold electrode was used in the electrochemical cell of the detector, and the settings were (Time in seconds/Volts): 0/0.1; 0.2/0.1; 0.4/0.1; 0.41/-2.0; 0.42/-2.0; 0.43/0.6; 0.44/-0.1; 0.5/-0.1. Run time was a total of 48 minutes, with the first 18 minutes for sample separation followed by a 15 minutes washing step with 200 mM NaOH, and a 15 minutes re-equilibration step with 90 mM

NaOH. Peak areas were integrated for galactinol, sucrose, raffinose, and stachyose. Carbohydrates were quantified based on standard curves generated for each carbohydrate. The content of galactinol, sucrose, raffinose, and stachyose is reported as the percent of dry seed weight, which can be converted to $g \cdot kg^{-1}$ by multiplying the percent of dry seed result by ten.

Data analysis.

Analysis of variance was conducted over four environments using PROC GLM of SAS (SAS Institute, 2004). To compare the stability of genotypes among environments for galactinol, sucrose, raffinose and stachyose in seed, and stability coefficients (a) in linear graphs were used as stability parameters. Stability coefficient was calculated from the regression of the mean of each contents of a line at an environment on an environmental index. The environmental index was the mean each content of all lines at an environment minus the mean each content of all lines averaged across the four environments. Genotypes having stability regression coefficients closest to zero are more stable, whereas those that deviate significantly from zero (either positive or negative) are considered less stable to changes across environments. PROC REG of SAS was used to calculate the two regression slopes.

Results

The objective of this study was to understand the environmental factor effects on sucrose and RFO in soybean seed. Soybean lines were developed and selected for combinations of different available alleles of the *RS2* and *RS3* genes (Table 3.1). We determined the amounts of galactinol, sucrose, raffinose, and stachyose extracted from randomly chosen 15 mature seeds from these soybean lines grown in different field environments. Based on the combination of alleles, soybean lines were divided into six different genotypic categories: those containing wild type (*RS2*/*RS3*), those with functional *RS2* gene and *rs3snp6* (*RS2*/*rs3snp6*), those with the *T107I* missense alleles of the *RS2* gene and functional *RS3* gene (*rs2T107I*/*RS3*), those with the *rs2T107I* missense alleles of the *RS2* gene and *rs3snp6* alleles (rs2*T107I*/RS3), those with the *W331−* alleles of the *RS2* gene and functional *RS3* gene (*rs2W331−*/*RS3*), and those with the *W331−* alleles of the *RS2* gene and *rs3snp6* alleles (*rs2W331−*/*rs3snp6*). Soybean lines were selected from segregating populations depending on alleles with gene-based markers. Williams 82, an inbred line with the *RS2*/*RS3* genotype, was used as check for this study. The tofu line 534545 was previously characterized as having more sucrose content than normal soybean, and it has the *RS2*/*rs3snp6* genotype. Fourteen soybean lines were tested to understand carbohydrate profiles in four environments.

Analysis of variance was conducted to identify location and year effects on each carbohydrate component measured, galactinol, sucrose, raffinose, and stachyose. The location (L) effect was significant for sucrose and raffinose content (Table 3.2). The year (Y) effect was significant for galactinol, sucrose and stachyose ($P < 0.01$). The L x Y interactions were significant for galactinol, raffinose and stachyose. Each of the measured carbohydrates in the

soybean lines were significant for genotype effects (G) ($P < 0.01$) which indicates that accumulation of carbohydrate component is genotype dependent. For the genotype (G) x environment interaction effects, the G x L interactions were significant for raffinose and stachyose ($P < 0.01$); the G x Y interactions were significant for galactinol and stachyose ($P < 0.01$); the G x L x Y interaction was significant only for stachyose (Table 3.2).

Overall, all of the carbohydrate components in soybean were mainly determined by genotype according to the combination of *RS2* and *RS3* alleles. Among the environmental factors, year effect was the highest factor to influence the contents of galactinol and sucrose (Table 3.2). For environment effects, the content of galactinol, sucrose, and stachyose in 2013 were significantly lower than in 2012 (Table 3.3). For location effects, the content of sucrose and raffinose were significantly different between two locations. The content of sucrose in soybean was highly contributed by environmental effects such as a year and a location factor (Table 3.2; Table 3.3).

Within genotypic classes, the variation of carbohydrates due to environmental factors and genotypic effects was demonstrated with a stacked bar graph of mean squares (Figure 3.1). The content of sucrose and stachyose was shown as those components are indicators to understand how each genotypic class was influenced by the factors. A larger portion indicated that soybean lines within genotypic classes with either sucrose or stachyose were more influenced by those environmental effects and genotype effects. For sucrose, the year effect was the major factor for observed variation in genotypic classes *RS2*/*RS3*, *RS2/rs2snp6*, *rs2T107I/RS3* and rs2*T107I/rs3snp6*, whereas in genotypic classes *rs2W331-/*RS3 and *rs2W331-/rs3snp6*, the major factor was the genotype effect (Figure 3.1.A). For stachyose, soybean lines with the genotypic

class *RS2*/*RS3*, year effect was a major factor. And year effect and year by location interaction effect were important in this study for *RS2*/*rs3snp6*. For the *rs2T107I/*RS3, *rs2T107I/rs3snp6* and *rs2W331-/*RS3 genotypic classes, several factors such as year, genotype and genotype by year interactions contributed to the variation in amount of stachyose. For the *rs2W331-/rs3snp6* genotypic class, there are many environmental factors involved in variation of the amount of stachyose (Figure 3.1.B).

Overall, year effect was the most important factor to determine the content of sucrose in soybean lines not having *rs2W331-* mutant alleles. Soybean lines with *rs2W331-* alleles were mostly affected by genotypic effect (Fig 3.1. A). Soybean lines with *RS2* alleles were influenced by year effect for the content of stachyose, whereas soybean lines with *rs2* mutant alleles were determined by genotype (Fig 3.1. B).

Mean values of each carbohydrate component of individual lines over four environments are shown in Figure 3.2. Significant differences were observed across the range of values for all four carbohydrate components among the lines. The mean content of galactinol in soybean lines ranged from 0.01 % to 0.82 % in dry seed matter (Figure 3.2.A). Within genotypic categories, the content of galactinol had no significant differences except *rs2W331-*/RS3 and *rs2W331-/rs3snp6*. The highest content of galactinol was a soybean line in the *W331-/rs3snp6* genotypic class (Figure 3.2.A). The mean of content of raffinose was less than 1.2 % in dry seed matter (Figure 3.2.B). Soybean lines containing mutant alleles of the *RS2* gene had a dramatically decreased level of raffinose content. Interestingly, a soybean line with *RS2*/*rs3snp6* was not significantly different from the content of raffinose in the wild type *RS2*/*RS3* genotypic class. Although the *RS2* and *RS3* genotype of the tofu lines is identical with the *RS2* and *rs3snp6*

alleles of two control lines, the content of raffinose in the tofu line was significantly lower than the control *RS2*/*rs3snp6* genotypes KB10-22#1538 and KB10-23#1634. Only lines in the genotypic class *rs2W331-/rs3snp6* were significantly different in the raffinose content within the class.

The mean sucrose content in soybean lines ranged from 5.0 % to 9.6 % in dry seed matter (Figure 3.2.C). Soybean lines with high concentration of sucrose were in the *rs2W331-*/*RS3* and *rs2W331-/rs3snp6* genotypic classes. The tofu line was significantly higher than the control *RS2*/*rs3snp6* lines. The amounts of sucrose were significantly different within genotypic classes such as *RS2*/*RS3*, *rs2W331-*/*RS3* and *rs2W331-/rs3snp6*. All soybean lines in *rs2T107I/rs3snp6*, *rs2W331-*/*RS3*, *rs2W331-/rs3snp6* genotypic classes and the tofu line were above 7 % of sucrose in this study (Figure 3.2.C). The mean content of stachyose in soybean lines ranged from 0.1 % to 6.2 % in dry seed matter (Figure 3.2.D). The soybean lines with *rs2W331-/rs3snp6* contained the least amounts of stachyose. Soybean lines containing mutant alleles of the *RS2* gene contained a decreased level of stachyose. The mean and variation for each carbohydrate component in each environment are shown in Figure 3.3. The genotypic classes with the highest level of sucrose content were soybean lines with *rs2W331-*/*RS3* and *rs2W331-/rs3snp6* genotypes in each environment, and soybean lines with *rs2W331-/rs3snp6* contained only a trace level of RFO in seeds in each environment.

Since galactinol, sucrose, raffinose, and stachyose are substrates and products in the biochemical pathway, it is likely that there are significant correlations among them from seed contents in different genotypic classes. According to statistical analysis, there was a significant negative correlation between raffinose and sucrose ($R = -0.48$, $P < 0.0001$) (Figure 3.4). Other

significant correlations were: negative between RFO and galactinol (R^2 = -0.73, P < 0.0001); positive between galactinol and sucrose (R^2 = 0.52, P < 0.0001), as well as positive between raffinose and stachyose (R = 0.85, P < 0.0001) (Figure 3.4).

The regressions of means from each carbohydrate from the six genotypic classes on the environmental index illustrates the environmental stability of the carbohydrate components. (Figure 3.5). The slope (a) of each of the genotypic classes for each carbohydrate was represented as variation among genotypic classes for environmental stability across four different environments. Genotypes with slopes closest to zero are more stable over the four environments. The range of slopes for galactinol was from 0.08 to 2.1 (Figure 3.5.A). Genotypes with *RS2*/*RS3* (a = 0.08) and *RS2*/*rs3snp6* (a = 0.09) were more stable than other genotypes for content of galactinol in seed. The range of slopes for sucrose were from 0.3 to 1.6 (Figure 3.5.B). Overall, soybean lines were relatively unstable for sucrose content over the four environments. Genotypic classes with *rs2W331-*/*rs3snp6* (a = 0.291) and *rs2W331-*/*RS3* (a = 0.481) had a high level of sucrose content and were comparatively stable. The range of slopes for raffinose was from 0.3 to 3.6 (Figure 3.5.C). Soybean lines with mutant alleles of the *RS2* gene were stable over four environments. In addition, those soybean lines had low levels of raffinose content. The range of slopes for stachyose were from 0.3 to 1.9 (Figure 3.5.D). The most stable and with the least amount of stachyose over environments was the genotypic class *rs2W331-/rs3snp6* (a = 0.283). Overall, the allele of *rs2W331-* is the most important allele for carbohydrate profile and environmental stability in soybean. Among all lines, those with *rs2W331-* mutant alleles of the *RS2* gene were the most stable over the four environments for sucrose, raffinose and stachyose.

The sum of galactinol, sucrose, raffinose, and stachyose (sum carbohydrates) in soybean lines with mutant alleles of the *RS2* gene was two or three percent lower than wild type (Figure 3.6). Linear regressions of protein and oil content on the amount of sum carbohydrates showed there was a negative correlation between sum carbohydrates and protein ($R = -0.40$, $P < 0.001$) and there was significant correlation between sum carbohydrates and oil ($R = 0.21$, $P = 0.0057$) (Figure 3.7).

A field emergence test was conducted in one year for soybean lines with altered seed carbohydrate profiles characterized in this work. The results indicated lines with the *rs2W331-/rs3snp6* genotype showed variable and low emergence percentage, and one of the two soybean lines with the *rs2W331-/RS3* genotype showed significantly low field emergence compared to the remaining lines.

Discussion

In the present study, we demonstrated the importance of both environment and genotype effects to carbohydrate profiles in soybean. The results indicated that galactinol and sucrose content were mainly affected by environmental factors, especially year effect, whereas genotype effect was a major factor to influence the accumulation of raffinose and stachyose contents in soybean. In addition, soybean lines containing the *RS2* gene with the *rs2W331-* alleles (Dierking and Bilyeu, 2008) have an increased sucrose content and a reduction in RFO over environments. Within *rs2W331-*/*RS3* and *rs2W331-*/*rs3snp6* genotypic classes, genotype effect was the more important factor for sucrose, whereas many factors were involved in the accumulation of stachyose content. Soybean lines with *rs2W331-* alleles of the *RS* gene were more stable for sucrose and RFO over four environments than other genotypic categories. Soybean lines with *rs2W331-* alleles of the RS2 gene are the best lines to use for animal and human utilization as those lines had high concentrations of sucrose and low levels of RFO. Soybean lines only having mutant alleles of *RS3* demonstrated that allele alone is not an important gene for carbohydrate profile in soybean. However, soybean lines with mutant alleles of the *RS2* gene (*rs2T107I* and *rs2W331-*) plus *rs3snp6* alleles showed better RFO phenotypes. This is the first time that the *rs2T107I*/*rs3snp6* genotype combination has been explored, and the results independently demonstrated the modifying effect of the *rs3snp6* alleles in combination with mutant *RS2* alleles in enhancing the reduced RFO seed phenotype.

The amount of sum carbohydrates in the genotypic class *rs2W331-/rs3snp6* had significantly two or three percent lower than wild type. In the present study, linear regressions of protein and oil content on the amount of sum carbohydrate indicated that there was a correlation

between total carbohydrate and protein and there was no significant correlation between total carbohydrate and oil. Another study concluded that there was a negative correlation between protein and total carbohydrate (sum of sucrose, raffinose and stachyose) (Wilcox and Shibles, 2001). Further studies are needed to determine where carbon is shifted in these soybean lines.

Energy from stored carbohydrates in seeds is thought to be necessary for germination (Peterbauer and Richter 2001; Bewley and Black 1994). Inhibition of RFO metabolism in soybean seeds did not decrease germination, suggesting that RFO metabolism is not a requirement for soybean germination (Dierking and Bilyeu 2009). In addition, there were no significant differences between soybean lines with low RFO phenotype and wild type for field emergence trait (Dierking and Bilyeu 2009; Neus et al., 2005). In the present study, soybean lines with the most altered carbohydrate phenotype showed low emergence percentage. Several other soybean lines with the *rs2W331-/rs3snp6* genotype have been produced in the field, and no defects in field emergence have been detected (data not shown). The *rs2W331-/rs3snp6* line with the significantly reduced emergence was from seed that matured just before frost (a MG V line produced in an MG IV environment), so seed quality and genetic background may have been issues. It is also possible that alleles are associated with low emergence in the field rather than low levels of RFO.

Another group reported seven soybean lines grown at three very different geographical locations (Kumar et al., 2010). Their results demonstrated that the sucrose content increased with cooler growing conditions. However, raffinose and stachyose values were genotype dependent (Kumar et al., 2010). Also, another study was conducted with soybean lines containing altered *RS2* alleles in five distinct locations in addition to five successive planting dates. They

demonstrated that late planting date lines or cool temperatures during pod filling were highly correlated with increased sucrose and decreased stachyose (Bilyeu and Wiebold, 2016). In the present study, the year effect determined accumulation of sucrose content in soybean seed, and raffinose and stachyose values were genotype dependent as well. We investigated the temperature in two locations over two years, and concluded that average temperatures in 2012 were lower in September than in 2013. Sucrose contents in 2012 were higher. The differences of temperatures between the two years was larger than between the two locations. Cool temperature during pod filling period may be an important factor to accumulate sucrose content in soybean seed. Further research will be required to better understand environmental factors in the carbohydrate profile in soybean grown in different locations over years.

Tables

Table 3.1. Nomenclature, soybean lines, RS2 and RS3 Allelic Status, Maturity information for Columbia and Novelty, Missouri

	Genotype		Maturity	
Name	RS2 gene	RS3 gene	Columbia MO	Novelty MO
Williams 82[a]	RS2[b]	RS3[c]	33[d]	23
SB-01	RS2	RS3	28	23
KB10-22#1583 a	RS2	*rs3snp6*	*30*	22
KB10-23#1634 a	RS2	*rs3snp6*	*20*	-[e]
KB10-22#1548 b	*rs2T107I*	RS3	31	26
397	*rs2T107I*	RS3	31	20
KB10-22#1562 0	*rs2T107I*	*rs3snp6*	*30*	24
KB10-22#1600 0	*rs2T107I*	*rs3snp6*	*30*	27
KB10-22#1608 0	*rs2T107I*	*rs3snp6*	*30*	25
KB10-23#1639 b	*rs2W331-*	RS3	36	-
KB07-15H?	*rs2W331-*	RS3	31	20
KB10-23#1681 0	*rs2W331-*	*rs3snp6*	*51*	-
934D201	*rs2W331-*	*rs3snp6*	*33*	-
534545 [f]	RS2	*rs3snp6*	*33*	30

[a]Williams 82, SB-01, 934D201 and 534545 are inbred lines which are cultivars or landraces. 397 is a mutation line from Williams 82. Others are the pedigree lines. [b]RS2 indicates functional raffinose synthase 2 gene which is similar to Williams 82 reference genome for Glyma06g18890. *rs2T107I* and *rs2W331-* indicate the mutant alleles. [c]RS3 indicates functional raffinose synthase 3 gene which is similar to reference genome for Glyma05g08950. rs3snp6 is the mutant allele of RS3. [d]Maturity date is indicated by the number of days after September 1st when plants reached full maturity (R8). [e]Missing data is indicated by "-". [f]534545 is the commercial line for making Tofu.

Table 3.2. Mean squares from analysis of variance of each carbohydrate composition for soybean genotype in two locations over two years

Source	df	Galactinol	Sucrose	Raffinose	Stachyose
Location (L)	1	0.01	16.24**	0.25*	0.16
Year (Y)	1	2.79**	91.18**	0.07	27.37**
L*Y	1	0.03*	0.02	0.35**	2.23*
Replication(L*Y)	8	0.01	1.34	0.01	0.32
Genotype (G)	13	0.69**	22.78**	2.73**	58.85**
G*L	13	0.01	1.84	0.14**	0.64**
G*Y	13	0.11**	2.50	0.06	2.38**
G*L*Y	13	0.01	2.50	0.04	0.46*
Error	98	0.01	1.60	0.04	0.25

*Significant at the0.05 probability level

**Significant at the 0.01 probability level

Table 3.3. Variation of contents for each carbohydrate in years and locations. The number with bold within each carbohydrate indicates significant difference at $p \leq 0.05$.

		Galactinol(%)	Sucrose(%)	Raffinose(%)	Stachyose(%)	Temp(°C)
Year	2012	0.41	8.02	0.41	3.36	18.4[a]
	2013	0.15	6.43	0.46	2.57	21.0
	$LSD_{0.05}$	0.03	0.39	0.07	0.16	
Location	Columbia	0.29	7.54	0.39	3.01	20.2[b]
	Novelty	0.28	6.93	0.47	2.92	19.2
	$LSD_{0.05}$	0.03	0.39	0.07	0.16	

a; Average of temperature of September in two locations in one year

b; Average of temperature of September in two years in one location

LSD; least significant difference

Figures

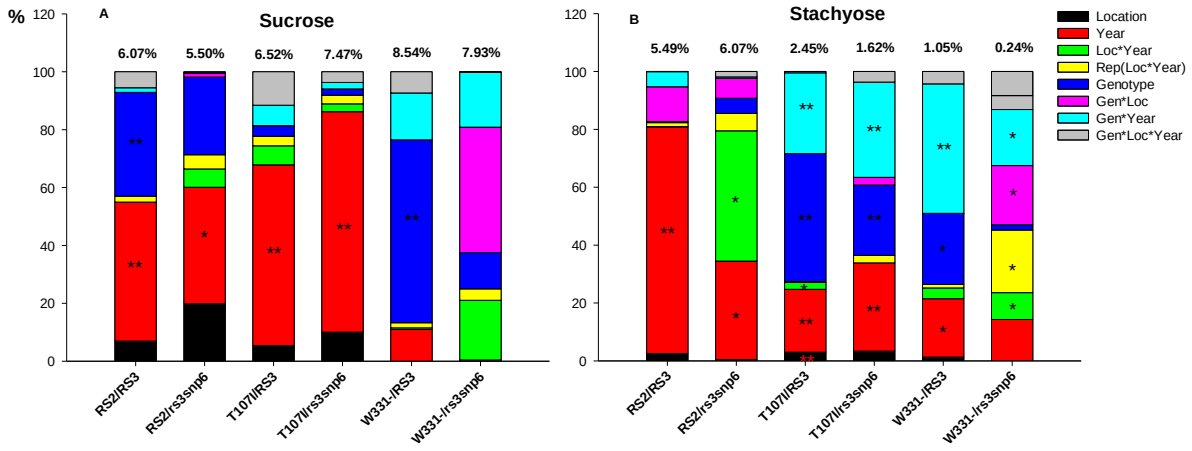


Figure 3.1. Comparison of relative contents of mean square of each environmental factors for soybean lines organized by RS2 and RS3 alleles combination into different classes. The values above the bar represent mean value of each genotype measured across years and locations.

Loc, Location; Rep, Replication; Gen, Genotype;

A, Galactinol; B, Sucrose; C, Raffinose; D, Stachyose.

*Significant at the 0.05 probability level within genotype

**Significant at the 0.01 probability level within genotype

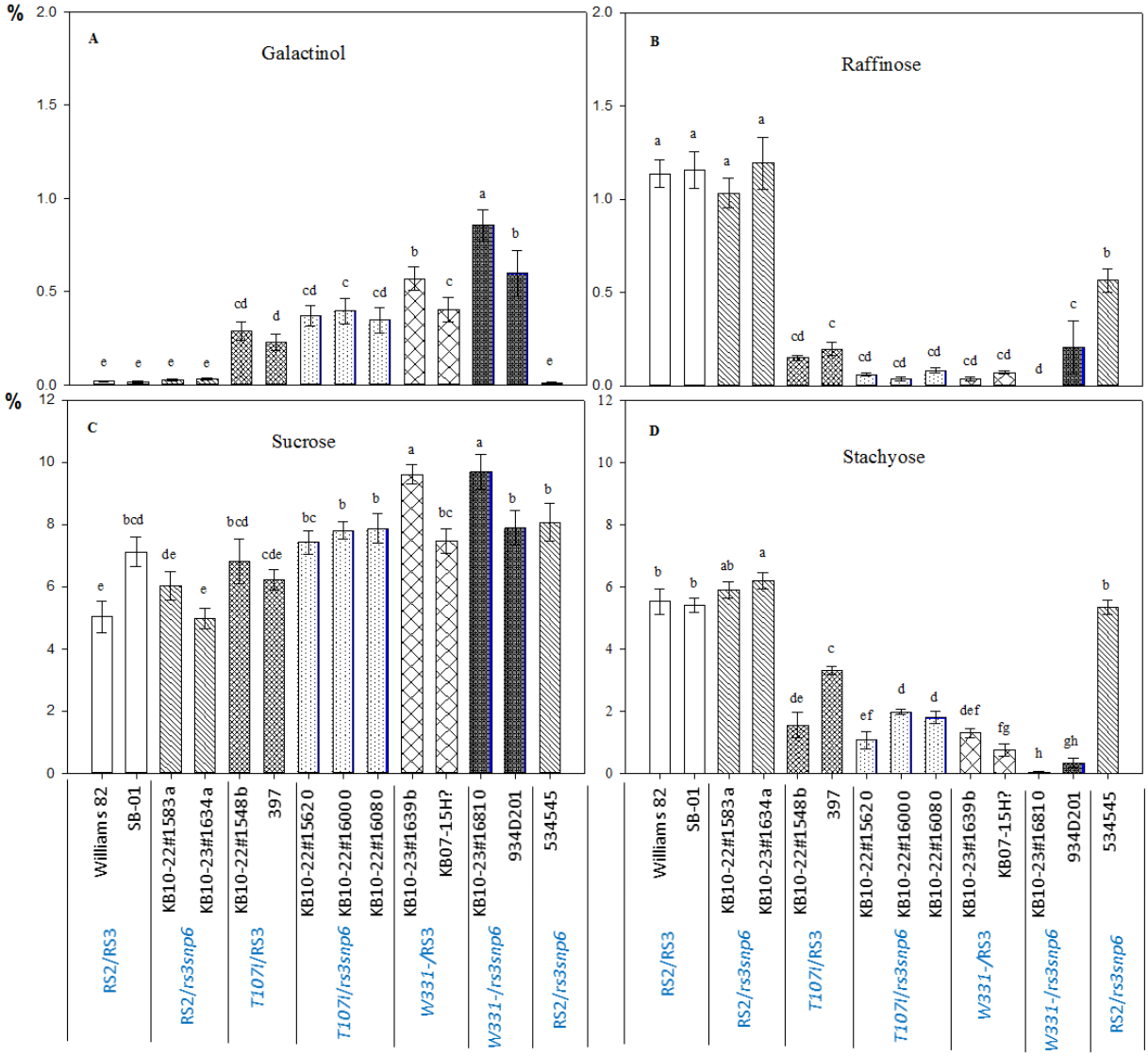


Figure 3.2. Contents of each carbohydrate of soybean lines averaged across two years and two locations. The bar represents standard error from mean value. The same letter within each carbohydrate indicates no significant difference at $p \leq 0.05$.

A, Galactinol; B, Raffinose; C, Sucrose; D, Stachyose.

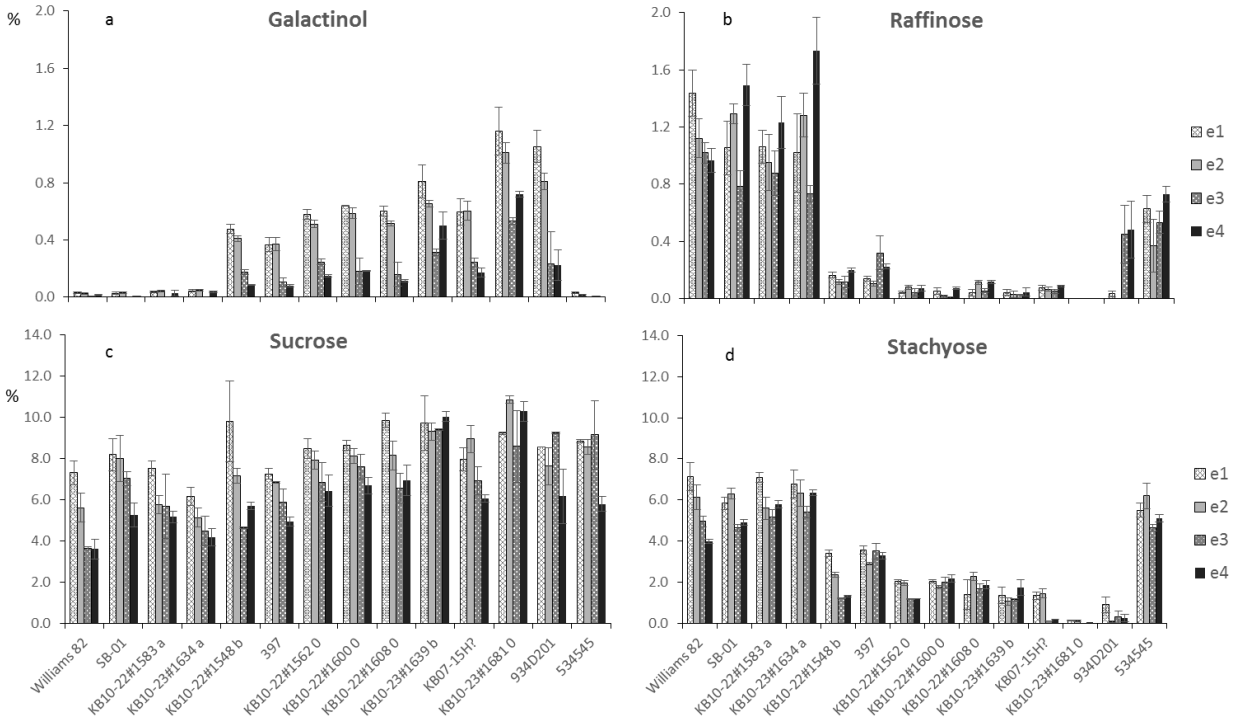


Figure 3.3. The mean values of galactinol (a), raffinose (b) sucrose (c) and stachyose (d) were shown over four different environments (E1, Columbia MO in 2012; E2, Novelty MO in 2012; E3, Columbia MO in 2013; E4, Novelty MO in 2013).

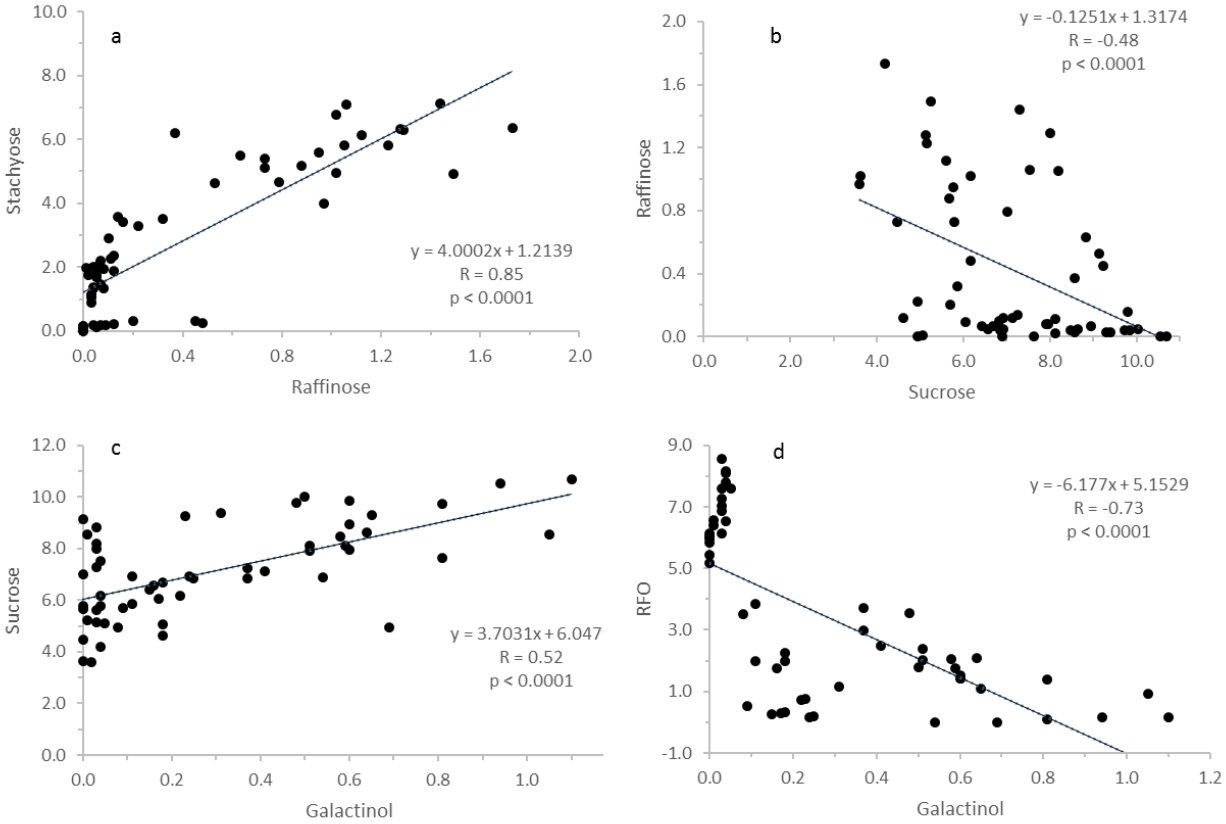


Figure 3.4. Correlation and regression of stachyose vs raffinose, raffinose vs sucrose, sucrose vs galactinol and RFO vs galactinol for all soybean lines over two years in two locations

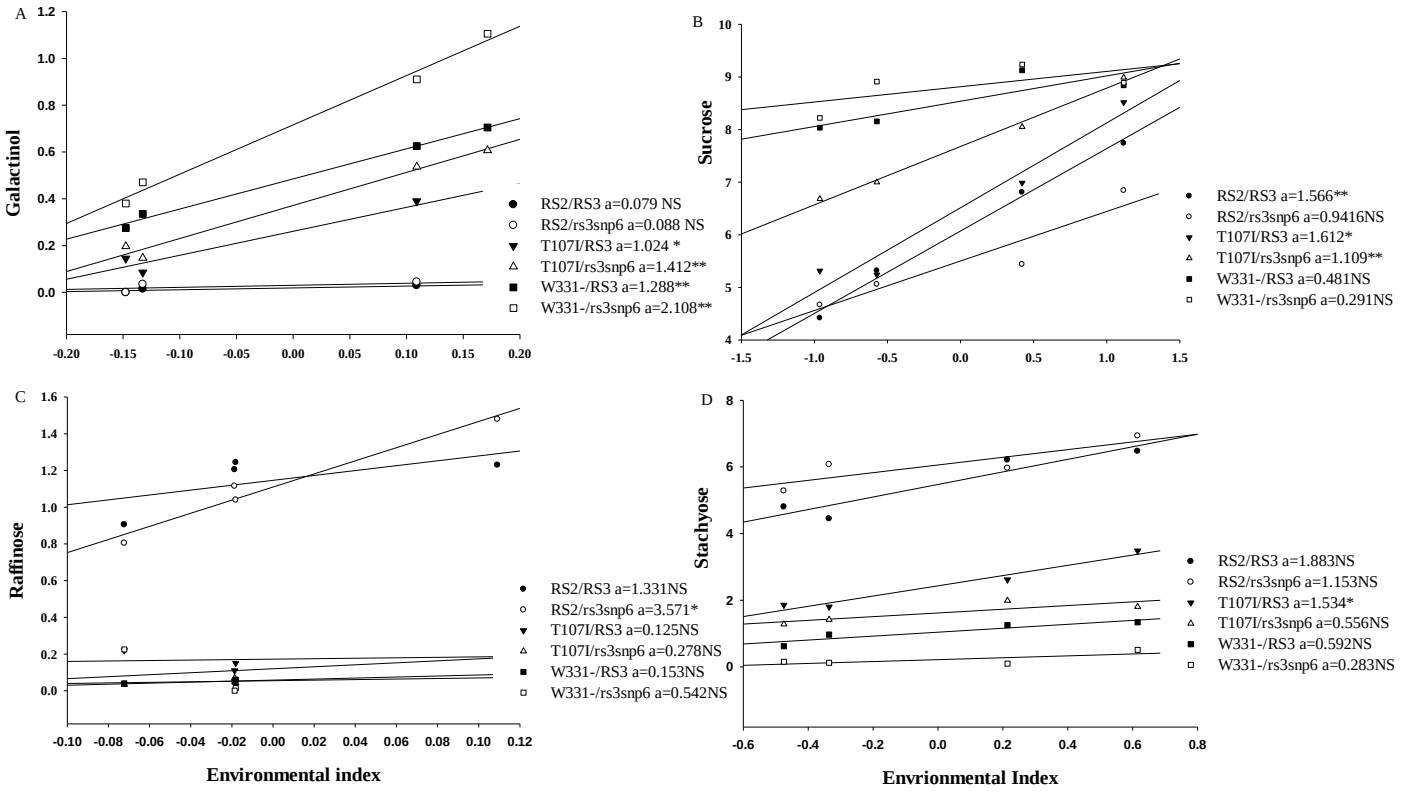


Figure 3.5. Regression graph of mean from each carbohydrate from each genotypic classes on environmental index. Environmental index was the mean of each carbohydrate of all lines at an environment minus the mean each carbohydrate content of all lines averaged across 4 environments. “a” means slope of each linear graph which is environmental stability.

*Significant at the 0.05 probability level;

**Significant at the 0.01 probability level;

NS, Not significant

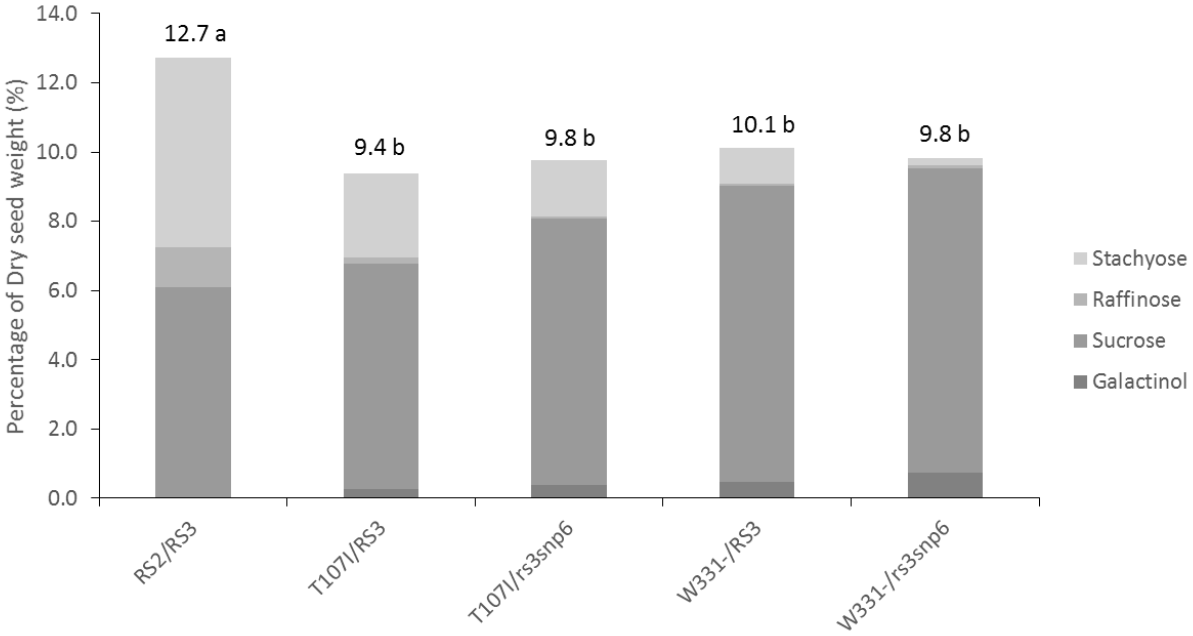


Figure 3.6. Comparison of overall carbohydrate profiles across four environments in each genotypic category. The value above the bar represents the sum of measured carbohydrates and the same letter means no significant differences ($P < 0.05$).

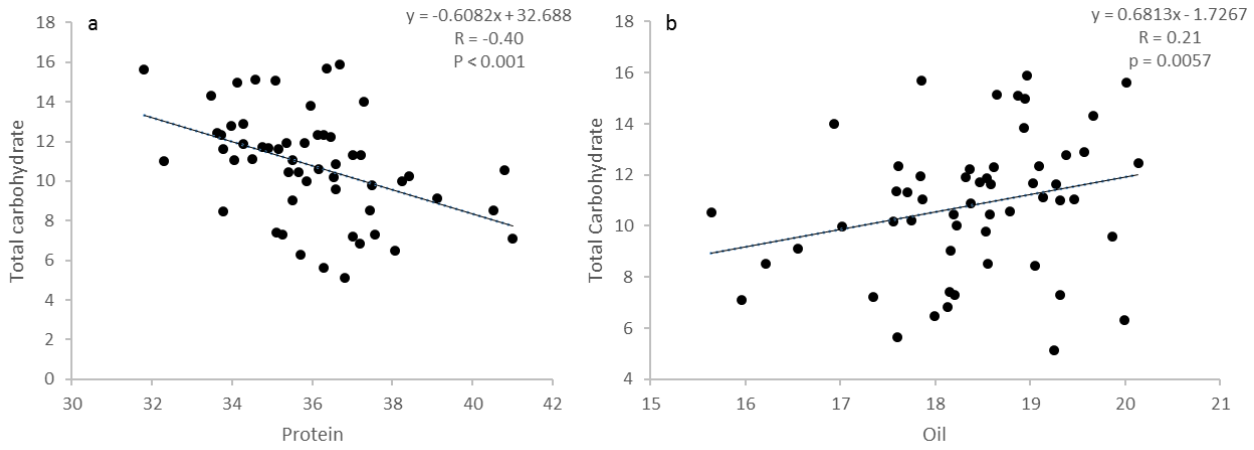


Figure 3.7. Correlation and regression of total carbohydrate vs protein and total carbohydrate vs oil for all soybean lines over two years in two locations

References

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.

Bewley, J.D., and M. Black. (1994) Seeds: Physiology of Development and Germination. Second ed. Plenum Press, New York.

Coon, C.N., K.L. Leske, O. Akavanichan, and T.K. Cheng. 1990. Eff ect of oligosaccharide-free soybean meal on true metabolizable energy andfi ber digestion in adult roosters. Poult. Sci. 69:787–793.

Colbert T, Till BJ, Tompa R, Reynolds S, Steine MN, Yeung AT, McCallum CM, Comai L, Henikoff S: High-throughput screening for induced point mutations. Plant Physiol 2001, 126(2):480-484.

Dierking, E.C. and Bilyeu, K.D. 2009 Raffinose and stachyose metabolism are not required for efficient soybean seed germination. Journal of Plant Physiology 166(12): 1329-1335.

Dierking, E.C. and Bilyeu, K.D. 2009 New sources of soybean seed meal and oil composition traits identified through tilling. BMC Plant Biol. 2009, 9, 89.

Dierking, E.C. and Bilyeu, K.D. 2008 Association of a soybean raffinose synthase gene with low raffinose and stachyose seed phenotype. Plant Genome 2008, 1 (2), 135−145.

Gitzelmann, R., & Auricchio, S. (1965). The handling of soya alpha-galactosides by a normal and a galactosemic child. Pediatrics, 36(2), 231-235.

Hagely KB, Palmquist D, Bilyeu KD (2013) Classification of Distinct Seed Carbohydrate Profiles in Soybean. Journal of Agricultural and Food Chemistry 61: 1105-1111.

Hitz, W. D.; Carlson, T. J.; Kerr, P. S.; Sebastian, S. A. Biochemical and molecular characterization of a mutation that confers a decreased raffinosaccharide and phytic acid phenotype on soybean seeds. Plant Physiol. 2002, 128 (2), 650−660.

Kerr, P.S., and S.A. Sebastian. 2000. Soybean products with improved carbohydrate composition and soybean plants. EP 0608359 B2, 2000.

Kumar, V. Rani, A.; Goyal, L.; Dixit, A. K.; Manjaya, J. G.; Dev, J.; Swamy, M. (2010) Sucrose and raffinose family oligosaccharides (RFOs) in soybean seeds as influenced by genotype and growing location. J. Agric. Food Chem., 58 (8), 5081−5085.

McCallum CM, Comai L, Greene EA, Henikoff S: Targeted screening for induced mutations. Nature Biotechnology 2000, 18:455-457.

Neus, J.D., Fehr, W.R. and Schnebly, S.R., 2005. Agronomic and seed characteristics of soybean with reduced raffinose and stachyose. Crop Science, 45(2), pp.589-592.

Peterbauer, T. and Richter, A. (2001). Biochemistry and physiology of raffinose family oligosaccharides and galactosyl cyclitols in seeds. Seed Science Research, 11(3), 185-198.

Price, K.R., J. Lewis, G.M. Wyatt, and G.R. Fenwick. 1988. Flatulence—causes, relation to diet and emedies. Nahrung 32:609–623.

Parsons, C.; Zhang, Y.; Araba, M. Nutritional evaluation of soybean meals varying in oligosaccharide content. Poult. Sci. 2000, 79 (8), 1127−1131.

Ruttloff , H., A. Täufel, W. Krause, H. Haenel, and K. Täufel. (1967). The intestinal enzymatic decompositon of galacto-oligosaccharides in the human and animal intestine, with

particular regard to Lactobacillus bifi dus. Part II: On the intestinal behaviour of lactulose. Nahrung 11:39–46.

SoyStats (2014) Available at http://www.soystats.com.

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

Suarez, F.L., J. Springfi eld, J.K. Furne, T.T. Lohrmann, P.S. Kerr, and M.D. Levitt. 1999. Gas production in humans ingesting a soybean fl our derived from beans naturally low in oligosaccharides. Am. J. Clin. Nutr. 69:135–139.

Smiricky, M. R.; Grieshop, C. M.; Albin, D. M.; Wubben, J. E.; Gabert, V. M.; Fahey, G. C. The influence of soy oligosaccharides on apparent and true ileal amino acid digestibilities and fecal consistency in growing pigs. J. Anim. Sci. 2002, 80 (9), 2433−2441.

Schillinger JA, Dierking EC, Bilyeu KD (2013) Soybeans having high germination rates and ultra-low raffinose and stachyose content. In USP Office, ed, USPTO, Vol 8,471,107. Schillinger Genetics, Inc. The United States of America as Represented by the Secretary of Agriculture, United States of America

Wilcox, J.R. and Shibles, R.M., 2001. Interrelationships among seed quality attributes in soybean. Crop Science, 41(1), pp.11-14.

Zuo, Y., G.C. Fahey, Jr., N.R. Merchen, and N.L. Bajjalieh. 1996. Digestion responses to low oligosaccharide soybean meal by ileally-cannulated dogs. J. Anim. Sci. 74:2441–2449.

Chapter 4

Environmental stability study of soybeans with modified carbohydrate profiles in soybean in maturity groups 0-V

Introduction

Soybean is one of the beneficial sources of animal meal due to the high protein content in the seed. Sixty-eight percent of world protein meal consumption was used from soybean seed (SoyStats 2015). Poultry consumes the majority of soymeal, using 55 % of U.S. soymeal, followed by swine and beef, which consume 22 % and 11 %, respectively (SoyStats 2015). Soluble carbohydrates in soybean mostly consist of sucrose and the raffinose family of oligosaccharides (RFO), raffinose and stachyose (Karr-Lilienthal et al. 2005; Wilson, 2004). Among the soluble carbohydrates in soybean, RFO are considered anti-nutritional factors as indigestible components for monogastric animals which cannot produce RFO-degrading enzymes in the digestive system (Gitzelmann and Auricchio, 1965; Ruttloff et al., 1967; Price et al., 1988; Sebastian et al., 2000). However, sucrose is the main beneficial component of soluble carbohydrates to increase metabolizable energy for animals. The modification of soluble carbohydrates in soybean seed which contain an increased sucrose content and a reduction of RFO is important for soybean meal industry and soybean producers.

The accumulation of soluble carbohydrate components in plants was associated with environmental effects. RFO accumulates during the process of cold acclimation (Gilmour et al., 2000; Koster et al., 1992). In leaves of tomato and Arabidopsis seedlings, raffinose synthase increased under environmental stress such as cold (Downie et al., 2003; Zuther et al., 2004). In addition, Wolf et al. (1982) studied the effect of low and high temperatures on carbohydrates in soybean seed. They concluded that raffinose levels were unchanged by low and high temperature, whereas sucrose and stachyose decreased at the highest temperature (Wolf et al., 1982). Thus,

accumulation of RFO is highly associated with temperature. Another group measured the carbohydrate profiles of seven soybean genotypes including Indian cultivars and exotic genotypes at three different geographical locations. They found that sucrose content was significantly higher at cooler locations, while raffinose and stachyose were genotype dependent at the three growing locations (Kumar et al., 2010). They suggested that soybean genotypes grown at cooler locations may be better able to produce soy food due to their carbohydrate characteristics (Kumar et al., 2010). Also, another study was conducted with soybean lines containing altered raffinose synthase 2 alleles in five distinct locations in addition to five successive planting dates. They demonstrated that late planting date lines or cool temperature during pod filling were highly correlated with increased sucrose and decreased stachyose (Bilyeu and Wiebold, 2016).

In addition, several studies investigated soybean with maturity date. Bellaloui et al. (2010) reported the effect of genotypic background of a maturity gene (E gene) on sugars. They indicated that there is no significant relationship with maturity date and raffinose, but there was a negative correlation between maturity date and sucrose, although NIR was used as the phenotyping method (Bellaloui et al., 2010). Soybean lines from four different genotypic classes based on raffinose synthase alleles including wild-type RFO, weak low RFO, low RFO and ultra-low RFO phenotypes were tested for carbohydrate profile on two different planting dates (Hagely et al., 2013). The results indicated that there was no consistent impact of maturity date on carbohydrate profile, but distinct genotypic classes were overall distinguishable (Hagely et al., 2013).

Soybean is a short-day and a photoperiod sensitive crop. The development of soybean is largely dependent on day-length, unlike many other crops that depend on temperature. Soybean will flower and transition to reproductive development stages from vegetative development stages when the night time lengthens beyond a critical dark period. Thus, soybean flowering is actually determined by the specific length of night time. Soybean varieties are sold under different maturity group designations to match the length of day and night time in the U.S. according to different latitudes. It is important for farmers to consider maturity of soybean to determine seed production and quality in a specific latitude in U.S. production environments. Soybeans that grow for a full season have a high quality and yield compared to very early matured soybeans or too late matured varieties.

There are several major soybean maturity categories in North America from Maturity Group (MG) 00, adapted to the border between U.S and Canada to MG VIII, adapted to the southern United States. Early maturity groups require a short dark period to transition to reproductive development, whereas late maturity groups need a long period of darkness to initiate flowering. When MG 00 varieties have been grown in the southern United States, there is not enough time for vegetative development before transition to reproductive stages, and production is sharply reduced. By contrast, late maturity groups grown in the northern U.S. may be affected by frost damage. Varieties in a given maturity group will usually perform as optimally full-season varieties within a band of similar latitude that is no wider than 100 to 150 miles from north to south.

The development of cultivars in a specific maturity group in U.S. requires effective use of maturity genes. Seven loci have been identified that influence the time to maturity. Alleles at

these loci act on photoperiod sensitivity. Using these loci can be difficult, and linked molecular markers could assist the soybean breeder in selecting for early maturity (Molnar et al., 2003). The maturity genes have been identified by classical methods. Several maturity loci designated as *E* loci have been found in soybean. *E1* is located on linkage group (LG) C2 at position 113.0 cM on the 2003 composite map (Song et al., 2004). *E7* is linked to *E1* and also resides on LG C2 (Molnar et al., 2003; Xia et al., 2012). E1 (Glyma06g23026) has been cloned and is a novel gene that appears to encode a transcription factor (Xia et al., 2012). There are several allelic variations in the *E1* gene including a single nucleotide polymorphism (SNP) at 44 base pair position with a missense mutation (arginine to threonine) (e1-as), a single base deletion at 49 base pair position with a premature stop codon at 124 base pair position (e1-fs), and the entire *E1* gene deletion, null allele (e1-nl) (Xia et al., 2012). *E2* is located at 136.3 cM on LG O (Glyma10g36600) which is an orthologue of the Arabidopsis GIGANTEA (GI) gene (Watanabe et al., 2011). A nonsense mutation was found at 1561 base pair when thymine base is substituted for adenine base (e2-ns) (Watanabe et al., 2011). *E3* is located on LG L (Glyma19g41210) (Molnar et al., 2003) which is a homolog of the photoreceptor phytochrome A (Liu et al., 2008). There are two functional *E3* alleles such as *E3-Ha* and *E3-Mi* (Langewisch et al., 2014). Both *E3* alleles have identical protein sequences. However, the *E3-Ha* allele, which is identical to the reference Williams 82 sequence, has an insertion in the third intron in comparison with the *E3-Mi* allele. There is no evidence to prove functionally difference between those *E*3 alleles (Liu et al., 2008). Having a 13-kb deletion in the *E3* gene, the mutant *e3-tr* allele does not contain the fourth exon (Liu et al., 2008). Another mutant *e3-ft3* allele has a missense mutation that glycine is substituted for arginine at 1050 position of the amino acid sequence (Liu et al., 2008).

In this study, we have developed soybean lines according to a combination of maturity genes that were predicted to explain a majority of the genetic variation for MGs 0 to V. These soybean lines contain different combinations of alleles of the raffinose synthase (RS) 2 and raffinose synthase 3 genes which control distinct seed carbohydrate profiles. The objective of this project was to understand the environmental stability of the carbohydrate profiles of soybean seeds with different *RS2* and *RS3* gene combinations adapted to different maturity groups across U.S. soybean production environments.

Materials and Methods

Soybean genotypes.

Soybean lines targeted to different maturity groups with different combinations of the *RS2* and *RS3* alleles were developed from breeding populations utilizing perfect molecular markers for the maturity genes *E1*, *E2*, and *E3* as well as perfect molecular markers for *RS2* and *RS3* alleles (Dierking and Bilyeu, 2008; Langewisch and Bilyeu, 2013; Xia et al., 2012; Watanabe et al., 2011). An *E* gene maturity model based on *E* gene haplotypes and phenotypic maturity data from the North American ancestral soybean lines as well as over 600 U.S. soybean cultivars was used to target different soybean maturity groups (Langewisch and Bilyeu, personal communication 2013). According to the model predictions for combinations of alleles at the *E1*, *E2* and *E3* loci, soybean lines were targeted from MG 0 to MG V (Table 4.1).

In addition, the goal was to select four different genotypic classes based on the combination of *RS2* and *RS3* mutant alleles in soybean lines for each maturity group. The mutant alleles in *RS2* included *W331-* and *T107I*, and the mutant allele in *RS3* was identified by a SNP in the first intron of the *RS3* gene (*rs3snp6)*. Modified RFO profile lines were organized by *RS2* and *RS3* alleles into four different classes: *T107I*/*RS3*, *T107I*/*rs3snp6*, *W331-*/*RS3* and *RS2W331-*/*rs3snp6*. Thirty-three soybean genotypes were tested in this study and Williams 82, grown in Novelty and Columbia, Missouri with wild type RFO content and functional alleles of *RS2* and *RS3*, was used as a check (PI 518671, United States, MG III) (Table 4.1).

Growth conditions.

The experiment was conducted at targeted locations from collaborators for some MG (Aaron Lorenz, University of Minnesota, Katy Martin Rainy, Purdue University in Indiana, and Pengyin Chen, University of Arkansas; the Grover Shannon location in the Portageville, Missouri location was lost), while the Novelty, Missouri and Columbia, Missouri location experiments were conducted by the Bilyeu laboratory (Figure 4.1). Ten seeds of each soybean line were planted and the experimental design was a randomized complete block design with three replications. Soybean lines with MG 0 and I were grown in the same location in Minnesota. MG II were planted in Indiana. Entire genotypes in MG III & IV were planted at the South Farm Research Center near Columbia MO (Central Missouri) and the Lee Greenley Jr. Memorial research center near Novelty, MO (Northern Missouri). Soybean lines in MG V were grown in Arkansas.

Seed carbohydrate phenotype determination

Soluble carbohydrates (galactinol, sucrose, raffinose, and stachyose) in soybean were determined by high performance ion chromatography with pulsed amperometric detection (PAD) employing a Dionex ICS-5000 with Electrochemical Detector (Thermo Scientific Dionex, Waltham, MA). Around 15 seeds from all genotypes were placed in -80 °C overnight and lyophilized to dryness prior to powdering. To phenotype, soluble carbohydrates in seeds were extracted by an ethanol extraction method. Briefly, 12.5 mg of sample was extracted with 1 mL of 50 % ethanol for 30 min at 70 °C with three times intermittent shaking in a 2 mL microcentrifuge tube. Then, samples were centrifuged 10 min at 16,000 x g. Approximately 700

μL of the supernatant was removed and stored at 4 °C before further experiments. A 50 μL aliquot of each sample was dried under vacuum and resuspended in 250 μL deionized water. Samples were placed in 96 well plate and automatically applied to the column with a 10 μL injection. Four soluble carbohydrates (galactinol, sucrose, raffinose, and stachyose) were separated on a Dionex Carbo Pac PA 10 analytical column (250 mm x 4 mm, 10 μm) connected to a Carbo Pac PA 10 guard column (50 nm x 4 nm). The mobile phase was 90 mM NaOH (blanketed with helium) with flow rate of 1.5 ml min^{-1}. A gold electrode was used in the electrochemical cell of the detector, and the settings were (Time in seconds/Volts): 0/0.1; 0.2/0.1; 0.4/0.1; 0.41/-2.0; 0.42/-2.0; 0.43/0.6; 0.44/-0.1; 0.5/-0.1. Run time was a total of 48 minutes, with the first 18 minutes for sample separation followed by a 15 minutes washing step with 200 mM NaOH, and a 15-minute re-equilibration step with 90 mM NaOH. Peak areas were integrated for galactinol, sucrose, raffinose, and stachyose. Carbohydrates were quantified based on standard curves generated for each carbohydrate. The content of galactinol, sucrose, raffinose, and stachyose is reported as the percent of dry seed weight, which can be converted to $g \cdot kg^{-1}$ by multiplying the percent of dry seed result by ten.

Data analysis.

Peak areas were integrated for galactinol, sucrose, raffinose, and stachyose. Carbohydrates were quantified on the basis of standard curves generated for each carbohydrate. Analysis of variance was conducted over four environments using PROC GLM of SAS (SAS Institute, 2004). To compare the stability of genotypes among environments for sucrose and stachyose content in seed, stability coefficients (a) in linear graph were used as stability

parameters. The stability coefficient was calculated from the regression of the mean of each of the carbohydrate contents of a soybean line at an environment on an environmental index. The environmental index was the mean of each content of all lines at an environment minus the mean of each content of all lines averaged across the four environments. Genotypes having stability regression coefficients closest to zero are more stable, whereas those that deviate significantly from zero (either positive or negative) are considered less stable to changes across environments. PROC REG of SAS was used to calculate the two regression slopes.

Results

Two analysis of variance (ANOVA) were conducted in this study. The first ANOVA was conducted between genotypic categories to identify genotype, location and G x L effects on each carbohydrate component measured (galactinol, sucrose, raffinose, and stachyose) (Table 4.2). In this case, each of the measured carbohydrates in the soybean lines were significant for genotype effect (G) and location effect (L) (Table 4.2). The three replications showed no significant differences for carbohydrate components besides sucrose content (Table 4.2). G x L effect showed significant effects for galactinol, raffinose and stachyose, but not sucrose (Table 4.2).

A second analysis was conducted within each genotypic class across different MG locations to identify effects for each carbohydrate component (Table 4.3). Sucrose content showed no significant difference for the *W331-/rs3snp6* genotypic category, whereas there were significant effects on galactinol and RFO content for that genotypic category (Table 4.3). For the *W331-/RS3* genotypic category, there were significant effects for all carbohydrate components (Table 4.3). The *T107I/rs3snp6* and *T107I/RS3* genotypic categories also showed significant effects for galactinol, sucrose and stachyose contents, except raffinose (Table 4.3).

The mean values of sucrose and stachyose content of the individual lines in the experiment are shown in Figure 4.2. The mean content of sucrose in soybean lines ranged from 5.9 % to 9.6 % in dry seed matter (Figure 4.2). The range of stachyose content was from 0.40 % to 6.9 %. The least significant difference (LSD) was 0.83 for sucrose content and 0.94 for stachyose content ($P < 0.05$). Williams 82, having wild phenotype of each carbohydrate component, was used as control and was only grown in Missouri. Seed of Williams 82 contained

5.9 % sucrose and 6.9 % stachyose. All soybean lines except one (KB13-18#1515-1 in the *rs2T107I/RS3* genotypic class) were significantly higher in sucrose content than Williams 82 (Figure 4.2). The highest sucrose content among soybean lines was in the *W331-/RS3* genotypic category grown in Arkansas (KB13-23 #1066-1). The content of stachyose in Williams 82 was significantly higher than all mutant soybean lines in this study (Figure 4.2).

Mean values of each carbohydrate component of individual lines in all environments were shown in Table 4.4. LSD was conducted within genotypic categories. For the *W331-/rs3snp6* genotypic category, soybean lines grown in different locations had no significant difference in the sucrose content. The raffinose content of soybean lines grown in MG 0 and MG I were significantly higher than other soybean lines in different MGs (Table 4.4.a). The stachyose content for MG 0, MG I and MG V were significantly higher than soybean lines in MG II and MG III & IV (Table 4.4.a). Among soybean lines with the *W331-/RS3* genotype, the sucrose content was not significantly different (Table 4.4.b). Soybean lines with the *W331-/RS3* genotype grown in MG 0 and MG II had significantly higher levels of stachyose than the soybean lines in MG I and KB13-23 #1066-1 in MG V (Table 4.4.b). In the *T107I/rs3snp6* genotypic category, KB13-18 #1539-705 in MG III & IV had significantly higher sucrose content than other soybean lines except KB13-18#1617-727 grown in MG I (Table 4.4.c). Soybean lines grown in MG I with *T107/rs3snp6* genotype showed statistically lower content of raffinose and stachyose than other soybean lines (Table 4.4.c). Soybean line KB13-20 #1657-9 in MG III & IV in the Columbia MO location only had statistically more sucrose content than other soybean lines with the *T107I*/*RS3* genotype. Soybean line, KB13-18 #1575-1 had the least amount of stachyose in the *T107I*/*RS3* genotypic category.

The regressions of means from each carbohydrate from the two genotypic classes with complete datasets on the environmental index illustrates the environmental stability of the carbohydrate components (Figure 4.3). The slope (a) of each of the genotypic classes for contents of sucrose and stachyose was represented as variation among genotypic classes for environmental stability across different environments. Genotypes with slopes closest to zero are more stable over different environments. Genotypes with *W331-/rs3snp6* (a = 0.73) were more stable than lines with *W331-/RS3* genotypes for sucrose. All mean sucrose values of lines with the *W331-/rs3snp6* genotype in different locations were higher than the *W331-/RS3* genotypes (Figure 4.3). However, the stachyose content showed the opposite result. The content of stachyose for lines with the *W331-/RS3* genotype (a = 0.0002) was more stable than *W331-/rs3snp6* genotypes. In addition, the mean of stachyose with *W331-/rs3snp6* genotypes in three locations had higher levels of stachyose content than *W331-/RS3* genotypes.

Discussion

This study was to extend the understanding of environmental stability of carbohydrate profile in soybean seed in multiple locations of soybean production areas in the U.S. Previous studies have investigated the stability of soluble carbohydrate profile with soybean lines having different combinations of *RS2* and *RS3* alleles with different planting dates in Missouri in the U.S (Hagely et al., 2013; Bilyeu and Wiebold, 2016). Because several studies reported that lower temperatures during pod filling of soybean correlated with an increased sucrose content and a reduction in RFO content (Kumar et al., 2010; Bilyeu and Wiebold, 2016), we developed soybean lines targeted to different MG that matured at the appropriate time for the environment using markers for maturity genes as well as selection for *RS2* and *RS3* alleles that have been shown to influence distinct carbohydrate profiles. By fixing the maturity alleles, soybean lines in this study matured at similar dates in each MG.

Across broader environments in this study using a variety of genetic backgrounds, most of the soybean lines in each targeted MG had more sucrose content and less RFO content than the control line with wild type *RS2* and *RS3* alleles; however, there were unexpected results for the accumulation of RFO content in the Minnesota and Arkansas locations. Because of this, the stability linear graph of stachyose content of *W331-/rs3snp6* lines was unstable in different MG locations. The results demonstrated that soybean lines with altered RFO phenotypes using mutant alleles of *RS2* and *RS3* were similar in certain regions of the Midwest U.S. such as Indiana and Missouri.

Sucrose was the most variable component of the measured soluble carbohydrates. Soybean lines in *W331-/rs3snp6* and *W331-/RS3* genotypic categories grown in Arkansas had the

highest levels of sucrose in this study. Temperatures during the soybean growing season in Arkansas were warmer than temperatures in Minnesota and Indiana. Soybean lines targeted to MG III and IV were planted in Central (Columbia) and Northern region (Novelty) of Missouri. Although most soybean lines in those two locations did not have significant differences for carbohydrate components within genotypic categories, there was a trend for soybean lines grown in Columbia to have slightly more sucrose than soybean lines in Novelty.

In addition, the Hou study in 2009 tested a total of 241 plant introductions from the USDA soybean germplasm collection from MG III, IV and V (Hou et al., 2009). Early MGs of soybean matured earlier than later MGs. Plant introductions with MG V had a significantly less amount of sucrose and RFO content than MG III. Also, another study supported this result that there were significant negative relationships between maturity (R8) and sucrose, as well as stachyose and combined sugars using near isogenic soybean lines with the same genotypic background with different combinations of genes for maturity (Bellaloui et al., 2010).

The source of mutant *RS2* alleles of *W331*- was the donor PI 200508 shown to have a low RFO phenotype and an increased sucrose concentration compared to wild-type plants (Kerr and Sebastian, 2002; Dierking and Bilyeu, 2008). PI 200508, MG I and determinate growth type contains maturity genotype *E1*, *e2* and *e3* alleles. The combination of maturity genes of MG I targeted soybean lines in this study was different from the one of PI 200508. From all these considerations, additional environmental factors influence the accumulation of sucrose content in soybean seed. These might be length of reproductive development stage, growth type, or maturity genes.

Soluble carbohydrate components were mostly affected by environmental factors and the previous study (Chapter 3) demonstrated year effect was the highest factor to influence sucrose content in soybean. Through stability studies, we can determine that threshold level of sucrose concentration in the soybean seed with any *RS2* mutant alleles is 7 % of the seed (dry weight basis). Above 7 % sucrose content in soybean may be considered high sucrose lines since the four environment stability study indicated lines with functional *RS2* and *RS3* alleles averaged 6 % sucrose (chapter 3). Soybean with ultra-low RFO content was defined as a less than 0.13 % of raffinose seed content and 1.6 % of stachyose content by dry weight of the total seed, respectively (Schillinger et al., 2013). However, here the ultra-low RFO lines contained less than 1 % of stachyose content in the dry seed weight. Low RFO soybean lines can be defined as less than 2 % of total amount of raffinose and stachyose in the dry seed weight. Sucrose content of soybean lines in the *W331-/rs3snp6* genotypic class were stable across locations in this study. Previous research (chapter 3) concluded that soybean lines with *W331-* alleles of *RS2* produced a seed carbohydrate phenotype classified as "low RFO" and "ultra-low RFO" that were more stable over four different environments than other genotypic classes. For stachyose content in the present study, soybean lines in *W331-/RS3* genotypic class were stable across locations, whereas soybean lines with *W331-* and *rs3snp6* alleles, classified as ultra-low RFO, were not stable across locations compared to the results in chapter 3, unexpectedly. Multiple years studies will be necessary to understand the accumulation of sucrose content in soybean seed, and other environmental factors will be considered for further research.

Tables

Table 4.1. Maturity genes allele, Locations, *RS2* and *RS3* allelic status, nomenclature and Maturity date.

Maturity Group	Location	*RS2* gene[a]	*RS3* gene[b]	Lines[c]	Maturity date[d]
MG0 (*e1-as e2 e3*)	Minnesota	*W331-*	*rs3snp6*	KB13-17 #1295-6	09/29
				KB13-17 #1295-9	09/27
		W331-	*RS3*	KB13-17 #1295-4	09/27
				KB13-17 #1226-7	10/02
MG I (*e1-as e2* E3)	Minnesota	*T107I*	*rs3snp6*	KB13-18 #1555-714	10/15
				KB13-18 #1617-727	10/5
		T107I	*RS3*	KB13-18 #1575-1	09/26
				KB13-18 #1617-1	10/02
		W331-	*RS3*	KB13-17 #1313-2	09/29
				KB13-17 #1313-5	09/28
		W331-	*rs3snp6*	KB12-31 #1008	10/07
MG II (*e1-as* E2 E3)	Indiana	*W331-*	*rs3snp6*	KB13-17 #1287-729	09/25
				KB13-19 #1396-745	09/22
		W331-	*RS3*	KB13-18 #1608-5	09/19
				KB13-20 #1702-7	09/20
		T107I	*RS3*	KB13-17 #1287-5	09/26
				KB13-19 #1442-8	09/27
MG III & IV (*e1-as* E2 E3)	Missouri (Novelty, Columbia)	*W331-*	*rs3snp6*	KB13-7 #1102-762	10/07
				KB13-17 #1331-790	09/28
		W331-	*RS3*	KB13-17 #1341-2	09/27
				KB13-17 #1233-6	09/25
		T107I	*rs3snp6*	KB13-18 #1539-705	10/04
				KB13-25 #1819-801	10/03
		T107I	*RS3*	KB13-18 #1515-1	09/30
				KB13-20 #1657-9	10/03
MG III	Missouri (Novelty, Columbia)	*RS2*	*RS3*	Williams 82 (PI 518671)	09/29
MG V (E1 E2 E3)	Arkansas	*W331-*	*rs3snp6*	KB13-23 #1064-815	10/02
				KB13-23 #1076-829	10/05
		W331-	*RS3*	KB13-23 #1066-1	10/02
				KB13-23 #1064-5	10/04
		T107I	*rs3snp6*	KB13-25 #1824-841	10/01
				KB13-25 #1824-844	10/04
		T107I	*RS3*	KB13-25 #1824-1	10/06
				KB13-25 #1824-3	09/29

[a]*RS2* indicates functional raffinose synthase 2 gene which is similar to Williams 82 reference genome for Glyma06g18890. *rs2T107I* and *rs2W331-* indicate the mutant alleles. [b]*RS3* indicates functional raffinose synthase 3 gene functional raffinose synthase 3 gene which is similar to reference genome for Glyma05g08950. *rs3snp6* is the mutant allele of *RS3*. [c]Williams 82 and the

pedigree lines. [d]Maturity date is an average of three replications when plants reached full maturity (R8). Red indicates soybean lines had a contamination issue and were excluded for further analysis.

Table 4.2. Mean squares from analysis of variance of each carbohydrates composition for soybean genotype in different location in 2015

		Galactinol (%)		Sucrose (%)		Raffinose (%)		Stachyose (%)	
	DF	MS	P-value	MS	P-value	MS	P-value	MS	P-value
Genotype (G)	3	0.25	<.0001	3.59	0.0006	0.16	0.0001	10.31	<.0001
Location (L)	5	0.20	<.0001	2.59	0.0010	0.09	0.0011	2.48	<.0001
Replication (L)	12	0.03	0.3753	1.16	0.0284	0.03	0.1585	0.53	0.1287
G*L	10	0.15	<.0001	0.49	0.5303	0.12	<.0001	4.88	<.0001

DF; degree of freedom

MS; mean square

Table 4.3. P value from analysis of variance of each carbohydrates composition within modified RFO genotypic classes

Within Genotype	P-value (ANOVA)			
	Galactinol	Sucrose	Raffinose	Stachyose
W331-/rs3snp6	0.001	0.336	0.028	0.001
W331-/*RS3*	<.0001	0.008	0.004	0.002
T107I/rs3snp6	0.048	0.035	0.082	<.0001
T107I/*RS3*	0.001	0.001	0.069	0.001

Table 4.4. Contents of each carbohydrate of soybean lines averaged with three replications.

a.

W331-/rs3snp6	Location	Line	Galactinol		Sucrose		Raffinose		Stachyose	
MG 0	Minnesota	KB13-17 #1295-6	0.24	E	8.37	A	0.63	A	3.95	A
MG I	Minnesota	KB12-31 #1008	0.47	DE	8.12	A	0.63	A	3.57	AB
MG II	Indiana	KB13-17 #1287-729	0.75	CD	8.08	A	0.08	B	0.44	D
MG II	Indiana	KB13-19 #1396-745	1.10	AB	8.25	A	0.07	B	0.47	D
MG III, IV	Novelty MO	KB13-7 #1102-762	1.12	A	8.36	A	0.07	B	0.47	D
MG III, IV	Columbia MO	KB13-7 #1102-762	0.82	ABC	9.21	A	0.03	B	0.46	D
MG V	Arkansas	KB13-23 #1064-815	0.77	BCD	9.39	A	0.13	B	1.83	CD
MG V	Arkansas	KB13-23 #1076-829	0.62	CD	8.62	A	0.15	B	2.18	BC

b.

W331-/*RS3*	Location	Line	Galactinol		Sucrose		Raffinose		Stachyose	
MG0	Minnesota	KB13-17 #1226-7	0.48	B	8.08	AB	0.19	A	1.71	A
MG I	Minnesota	KB13-17 #1313-2	0.68	B	7.85	AB	0.1	BC	0.82	B
MG I	Minnesota	KB13-17 #1313-5	1.09	A	6.89	B	0.12	ABC	0.95	B
MG II	Indiana	KB13-20 #1702-7	0.59	B	7.73	AB	0.17	AB	1.79	A
MG III, IV	Novelty MO	KB13-17 #1341-2	0.68	B	7.29	B	0.13	ABC	1.31	AB
MG III, IV	Columbia MO	KB13-17 #1341-2	0.70	B	7.97	AB	0.12	ABC	1.35	AB
MG V	Arkansas	KB13-23 #1066-1	0.71	B	9.44	AB	0.08	C	0.85	B
MG V	Arkansas	KB13-23 #1064-5	0.71	B	8.36	AB	0.10	BC	1.05	AB

c.

T107I/rs3snp6	Location	Line	Galactinol		Sucrose		Raffinose		Stachyose	
MG I	Minnesota	KB13-18 #1555-714	0.83	A	7.27	B	0.09	B	0.85	B
MG I	Minnesota	KB13-18 #1617-727	0.69	ABC	7.97	AB	0.09	B	0.82	B
MG III, IV	Novelty MO	KB13-18 #1539-705	0.75	AB	8.82	A	0.16	AB	1.91	A
MG III, IV	Novelty MO	KB13-25 #1819-801	0.59	BC	7.31	B	0.18	AB	2.30	A
MG III, IV	Columbia MO	KB13-18 #1539-705	0.62	BC	9.30	A	0.16	AB	1.90	A
MG III, IV	Columbia MO	KB13-25 #1819-801	0.50	C	7.38	B	0.31	A	2.37	A

d.

T107I/*RS3*	Location	Line	Galactinol		Sucrose		Raffinose		Stachyose	
MG I	Minnesota	KB13-18 #1575-1	0.92	A	7.35	B	0.11	B	0.97	C
MG I	Minnesota	KB13-18 #1617-1	0.69	AB	7.66	B	0.28	AB	2.21	B
MG II	Indiana	KB13-17 #1287-5	0.47	BC	7.47	B	0.23	AB	2.85	AB
MG III, IV	Novelty MO	KB13-18 #1515-1	0.37	C	6.26	C	0.43	A	3.78	A
MG III, IV	Novelty MO	KB13-20 #1657-9	0.46	BC	7.08	B	0.43	A	3.36	A
MG III, IV	Columbia MO	KB13-18 #1515-1	0.42	C	7.06	B	0.34	AB	3.35	A
MG III, IV	Columbia MO	KB13-20 #1657-9	0.36	C	8.50	A	0.42	A	3.32	A

The same letter within each carbohydrate indicates no significant difference at $p \leq 0.05$. Red indicates unusual accumulation of carbohydrate component in genotypic class.

a, W331-/rs3snp6; b, W331-/*RS3*; c, T107I/rs3snp6; d, T107I/*RS3*.

Figures

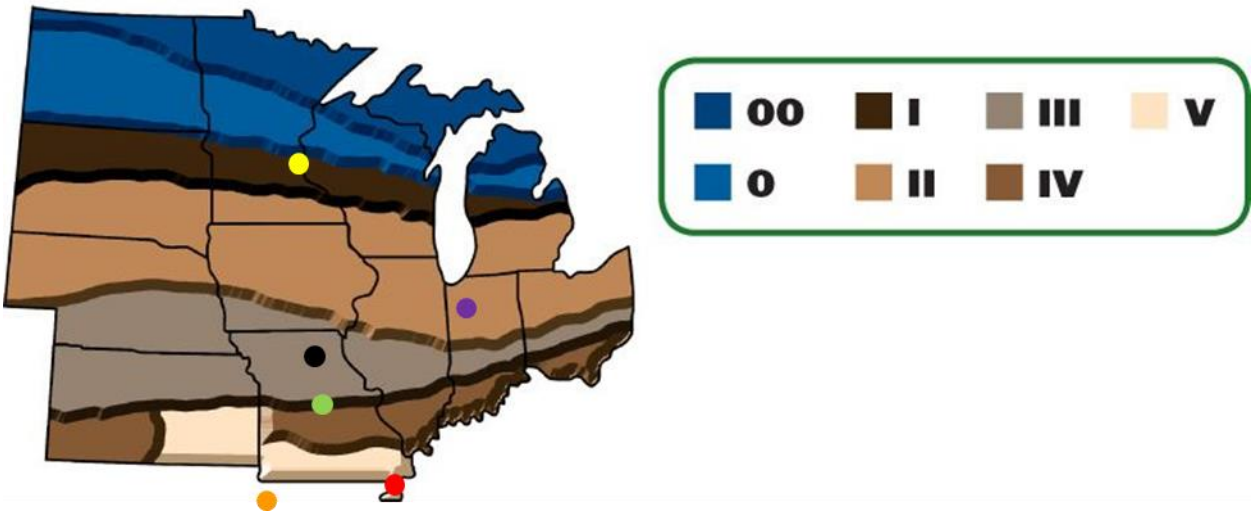
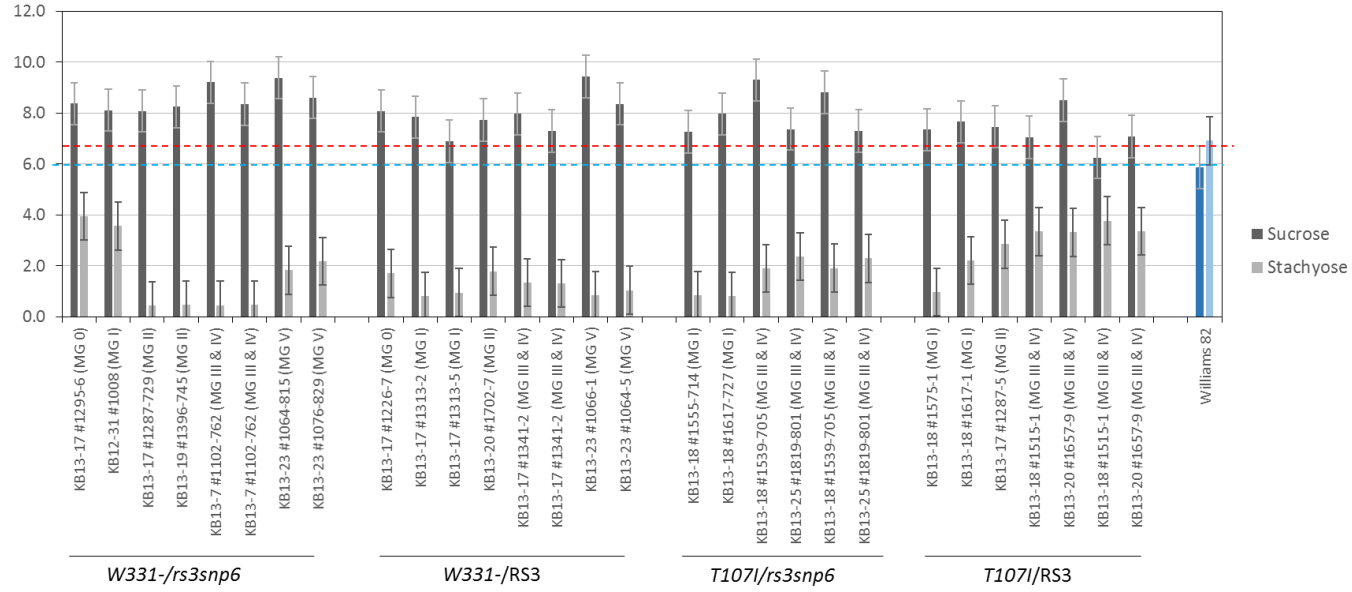


Figure 4.1. Map of Maturity Groups and study locations in the U.S. Colored dots represent approximate study site locations in Minnesota (MG 0 and I, yellow), Indiana (MG II, purple), Missouri (MG III, black; MG IV, light green; and MG V, red), and Arkansas (MG V, orange). Maturity Group map image modified from http://www.nutechseed.com/index.php?pkey=150

a



b

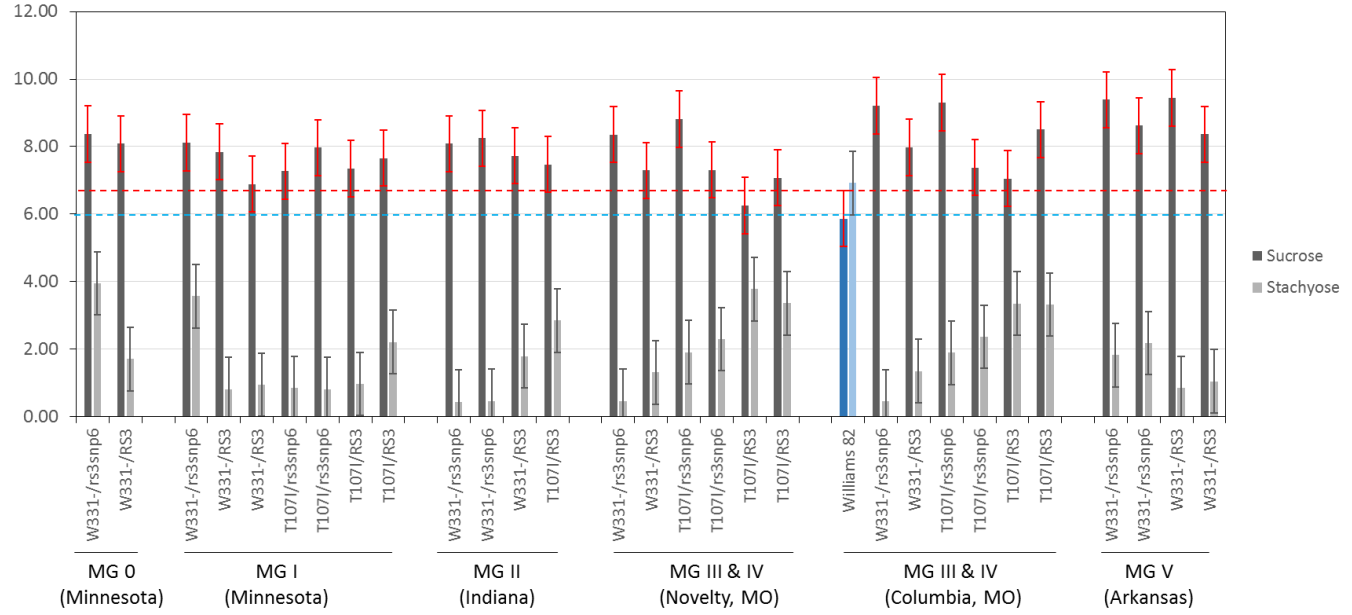


Figure 4.2. Contents of sucrose and stachyose of soybean lines averaged with three replications. The bar represents least significant difference (LSD) from mean value at $p \leq 0.05$. Red dashed line represents upper LSD range of the sucrose content of Williams 82. Blue dashed line

represents lower LSD range of the stachyose of Williams 82.

a. ordered by genotype b. ordered by MGs

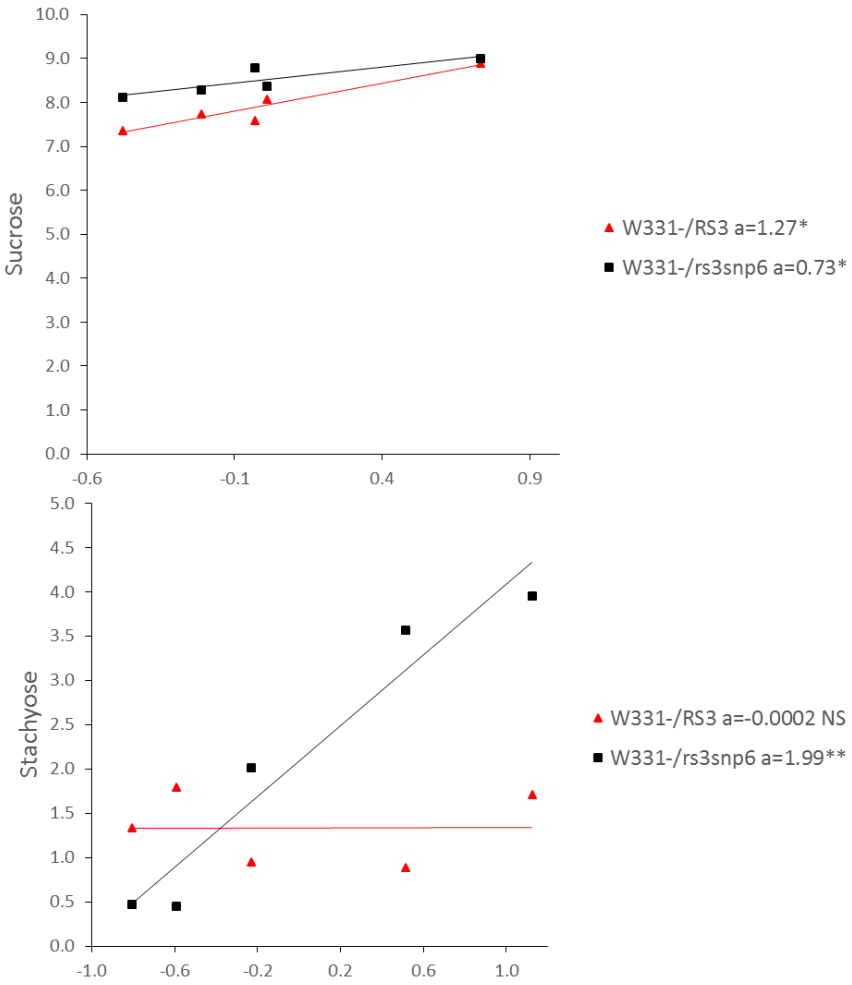


Figure 4.3. Regression graph of mean from sucrose and stachyose from two genotypic classes on environmental index. Environmental index was the mean of each carbohydrate of all lines at an environment minus the mean each carbohydrate content of all lines averaged across all environments. “a” means slope of each linear graph which is environmental stability. *Significant at the 0.05 probability level; **Significant at the 0.01 probability level; NS, Not significant

References

Bellaloui, N., Smith, J. R., Gillen, A. M., & Ray, J. D. (2010). Effect of Maturity on Seed Sugars as Measured on Near-Isogenic Soybean () Lines. Crop science, 50(5), 1978-1987.

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.

Downie B., Gurusinghe S., Dahal P., Thacker R.R., Snyder J.C., Nonogaki H., Yim K., Fukanaga K., Alvarado V., Bradford K.J. (2003) Expression of a GALACTINOL SYNTHASE gene in tomato seeds is up-regulated before maturation desiccation and again after imbibition whenever radicle protrusion is prevented. Plant Physiol 131: 1347–1359

Dierking, E. C., & Bilyeu, K. D. (2009). New sources of soybean seed meal and oil composition traits identified through TILLING. BMC plant biology, 9(1), 89.

Dierking, E. C., & Bilyeu, K. D. (2008). Association of a soybean raffinose synthase gene with low raffinose and stachyose seed phenotype. The Plant Genome, 1(2), 135-145

Franklin KA, Quail PH (2010) Phytochrome functions in Arabidopsis development. J Exp Bot 61:11–24.

Gitzelmann, R., & Auricchio, S. (1965). The handling of soya alpha-galactosides by a normal and a galactosemic child. Pediatrics, 36(2), 231-235.

Gilmour, S. J., Sebolt, A. M., Salazar, M. P., Everard, J. D., & Thomashow, M. F. (2000). Overexpression of the Arabidopsis CBF3transcriptional activator mimics multiple

biochemical changes associated with cold acclimation. Plant physiology, 124(4), 1854-1865.

Hagely, K. B., Palmquist, D., & Bilyeu, K. D. (2013). Classification of distinct seed carbohydrate profiles in soybean. Journal of agricultural and food chemistry, 61(5), 1105-1111.

Hou, A., Chen, P., Alloatti, J., Li, D., Mozzoni, L., Zhang, B. and Shi, A., (2009). Genetic variability of seed sugar content in worldwide soybean germplasm collections. Crop Science, 49(3), pp.903-912.

Karr-Lilienthal LK, Grieshop CM, Spears JK, Fahey Jr. GC (2005) Amino acid, carbohydrate, and fat composition of soybean meals prepared at 55 commercial U.S. soybean processing plants. J Agric Food Chem 53:2146-2150

Kumar, V., Rani, A., Goyal, L., Dixit, A. K., Manjaya, J. G., Dev, J., & Swamy, M. (2010). Sucrose and raffinose family oligosaccharides (RFOs) in soybean seeds as influenced by genotype and growing location. Journal of agricultural and food chemistry, 58(8), 5081-5085.

Koster, K. L., Lynch, D. V. (1992). Solute accumulation and compartmentation during the cold acclimation of Puma rye. Plant Physiology, 98(1): 108-113.

Liu B, Kanazawa A, Matsumura H, Takahashi R, Harada K, et al. (2008) Genetic redundancy in soybean photoresponses associated with duplication of the phytochrome A gene. Genetics 180: 995–1007.

Langewisch, T., Zhang, H., Vincent, R., Joshi, T., Xu, D. and Bilyeu, K., (2014). Major soybean maturity gene haplotypes revealed by SNPViz analysis of 72 sequenced soybean genomes. PloS one, 9(4), p.e94150.

Molnar, S. J., Rai, S., Charette, M., & Cober, E. R. (2003). Simple sequence repeat (SSR) markers linked to E1, E3, E4, and E7 maturity genes in soybean. Genome, 46(6), 1024-1036.

Price, K.R., J. Lewis, G.M. Wyatt, and G.R. Fenwick. (1988). Flatulence—causes, relation to diet and emedies. Nahrung 32:609–623.

Ruttloff , H., A. Täufel, W. Krause, H. Haenel, and K. Täufel. (1967). The intestinal enzymatic decompositon of galacto-oligosaccharides in the human and animal intestine, with

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

Song, Q., L. Marek, R. Shoemaker, K. Lark, V. Concibido, X. Delannay, J. Specht, and P. Cregan. (2004). A new integrated genetic linkage map of the soybean. Theor. Appl. Genet. 109:122–128.

particular regard to Lactobacillus bifi dus. Part II: On the intestinal behaviour of lactulose. Nahrung 11:39–46.

SoyStats (2014) Available at http://www.soystats.com.

Wilson RF (2004) Seed composition. In: Boerma HR, Specht JE (eds) Soybeans: improvement, production and uses. ASA, CSSA and SSSA, Madison, WI, pp 621-677

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

Watanabe, S., Xia, Z., Hideshima, R., Tsubokura, Y., Sato, S., Yamanaka, N., Takahashi, R., Anai, T., Tabata, S., Kitamura, K. and Harada, K., (2011). A map-based cloning strategy employing a residual heterozygous line reveals that the GIGANTEA gene is involved in soybean maturity and flowering. Genetics, 188(2), pp.395-407.

Xia Z,Watanabe S,YamadaT, et al. (2012). Positional cloning and characterization reveal the molecular basis for soybean maturity locus E1 that regulates photoperiodic flowering. Proceedings of the National Academy Sciences, USA 109: E2155–E2164.

Zuther E., Buchel K., Hundertmark M., Stitt M., Hincha D.K., Heyer A.G. (2004) The role of raffinose in the cold acclimation response of Arabidopsis thaliana. FEBS Lett 576: 169–173

Chapter 5

Association mapping of QTLs for high sucrose content in soybean using a genotyping by sequencing (GBS) approach

Introduction

Soybean is a one of the most important sources of protein and oil in the world. Soluble carbohydrates in soybean mostly consist of sucrose and the raffinose family of oligosaccharides (RFO), raffinose and stachyose (Karr-Lilienthal et al. 2005; Wilson 2004). Sucrose affects the quality of soybean meal and soybean food. Among soluble carbohydrate components, sucrose is the only beneficial component for animals, and it contributes to the sweetness of soymilk and tofu for human consumption. The modification of soluble carbohydrates in soybean seed which contain an increased sucrose content is important to achieve value for the soybean industry and farmers.

In soluble carbohydrates, there are sucrose and oligosaccharides that total about 12 % of the dry seed weight (Karr-Lilienthal et al. 2005; Wilson 2004). The amount of sucrose in soybean seed ranges from 1.5 to 10.2% of total dry matter (Hymowitz and Collins, 1974). Accumulation of sucrose is ceased when the color of seeds is changed from green to yellow during seed maturation (Obendorf et al., 1998).

Sucrose is the most variable component among soluble carbohydrates in soybean seed by environmental effects, especially cooling temperature during pod filing. There was a study that measured the carbohydrate profiles of seven soybean genotypes at three different geographical locations. They found that sucrose was significantly higher at cooler locations (Kumar et al., 2010). They suggested that soybean genotypes grown at cooler locations may be better able to produce soy food due to their carbohydrates' characteristics (Kumar et al., 2010). Also, another study was conducted with soybean lines containing altered raffinose synthase 2 alleles in five distinct locations in addition to five successive planting dates. They demonstrated

that late planting date lines were highly correlated with increased sucrose (Bilyeu and Wiebold, 2016). In addition, Wolf et al. (1982) studied the effect of low and high temperatures on carbohydrates in soybean seed. They concluded that raffinose levels were unchanged by low and high temperature, whereas sucrose decreased at the highest temperature (Wolf et al., 1982).

An understanding of the genetic factors that contribute to sucrose content in soybean could help breeders accelerate the development of varieties for food quality purposes. Maughan et al (2000) analyzed 149 F_2 individuals from a cross of *G. max* and *G. soja* with 178 polymorphic genetic markers, including RFLPs, SSRs, and RAPDs. They concluded that seventeen marker loci, mapping to seven different genomic regions (Chromosome 5, 7, 8, 13, 15, 19 and 20), were significantly associated with sucrose variation (Maughan et al., 2000). Also, they reported that a comparison of mapping results for sucrose and those for protein and oil in soybean suggested that seed composition traits were inherited as clusters of linked loci (Maughan et al., 2000). Another group also tested 115 F_{10} lines developed by single seed descent (SSD) for molecular markers to identify the genomic regions significantly associated with sucrose content (Kim et al., 2005). They found that QTLs for sucrose content are located at chromosome 11 (satt197), chromosome 2 (satt546), and chromosome 19 (satt523 and satt278) (Kim et al., 2005). Two of them (chromosome 2 and 19) were associated with oligosaccharide content. Two flanking markers on chromosome 19 tagged the major QTL for sucrose, which showed 21.4 % phenotypic variation for sucrose. Another RIL population was used to map QTLs on chromosome 12 and 16 for sucrose content (Kim et al., 2006). For oligosaccharide contents, four QTLs (chromosome 6, 12, 16 and 19) were detected. Phenotypic variations for each tagging markers were less than 10 % for sucrose content. Although the two RIL populations in the two

studies shared the high sucrose donor soybean line “Keunolkong”, the QTLs for sucrose in the two mapping populations did not overlap. Zeng et al. (2014) have recently reported on sucrose content in PI 243545 which is a germplasm introduced from Japan (USDA-ARS, 2009). Three QTLs were mapped on chromosome 5, 9 and 16 which have 46 %, 10 % and 8 % of phenotypic variation for sucrose content.

In this study, we have developed a mapping population between two soybean lines with similary maturity but different genetic mechanisms to achieve high sucrose content. The objective of this project was to identify locations on the soybean genome associated with sucrose content using the genotyping-by-sequencing (GBS) method.

Materials and Methods

Plant material and population development

In this study, a mapping population was developed from the cross between food-grade soybean cultivar 534545 (Brown, 2006) and soybean line, SGUL (Schillinger et al., 2013) at South Farm Research Center near Columbia, MO. F_1 seeds were sent to Costa Rica for generation advancement to develop Recombinant Inbred Lines (RILs). In the summer of 2015, 90 $F_{4:5}$ RILs were planted in three locations; South Farm Research Center near Columbia, MO (Central Missouri), Lee Greenley Jr. Memorial Research Center near Novelty, MO (Northern Missouri) and affiliated experimental fields of Kyungpook National University (Gunwi, South Korea). Ten seeds of each of the soybean lines were planted by hand into 91 cm plots with 30 cm spacing.

The determination of sucrose by enzymatic assay

Sucrose content in soybean was analyzed by colorimetric assay using invertase (β-fructosidase) and GOPOD enzyme containing glucose oxidase and peroxidase and 4-aminoantipyrine (Megazyme) (Teixeira et al., 2012). Around 15 seeds from all genotypes were placed in -80 °C overnight and lyophilized to dryness prior to powdering. To phenotype, soluble carbohydrates in seed were extracted by an ethanol extraction method. Briefly, 12.5 mg of sample was extracted with 1 mL of 50 % ethanol for 30 min at 70 °C with three times intermittent shaking in a 2 mL microcentrifuge tube. Then, samples were centrifuged 10 min at 16,000 x g. Approximately 700 µL of the supernatant was removed and stored at 4 °C before further experiments. 100 µL was transferred to 96-well PCR plate with 100 µL deionized water

(50:50 dilution) and mixed by pipetting. 10 µL of the dilution was placed into the well of a new plate containing 96 conical wells with 10 µL of invertase solution. In this step, each sample was duplicated in the 96-well plate. The plate was then covered by a lid and incubated in the oven for 20 min at 50 °C. After this, 20 µL of standard D-glucose solution was added to the plate. A solution of glucose standard was duplicated and used from a stock solution (1.0 mg/mL) in the kit and prepared at concentrations of 0 %, 25 %, 50 %, 75 % and 100 % of 0.005 µg/µL glucose solution. 150 µL of GOPOD was added to every well in the plate. The plate was then covered and incubated in the oven for 20 min at 50 °C again. 125 µL of each sample and standard solution was transferred to 96-well spectrophotometer plate and read by a spectrophotometer at 510 nm. The absorbance values of samples were calculated to percent of sucrose content in the seed.

DNA extraction and genotyping by sequencing

Young trifoliate leaves were collected from $F_{4:5}$ plants from RILs in the 2015 summer. Before starting DNA extraction, leaves of each line were freeze-dried and ground into a powder. 20 mg of lyophilized leaf tissue from each sample was placed into tubes and each sample of DNA was isolated using the DNeasy Plant Mini kit (Qiagen, Valencia, CA, USA). Quantification of genomic DNA was conducted by QuantiFluor dsDNA System (Promega, Madison, WI, USA) and DNA quality determined by running in 1.5% agarose gel. 30 µL of genomic DNA at 100 ng/µL from 90 RILs including parental lines were sent to the Institute for Genomic Diversity (IGD) at Cornell University. Genotyping-by-Sequencing (GBS) libraries were prepared as previously described (Elshire et al, 2011, Swarts et al, 2014) using the ApeKI

enzyme and appropriate Illumina adapters. The IGD staff performed the library construction, read mapping, and downstream SNP calls.

Association mapping

A total of 139,006 SNPs were identified. After those were filtered using the TASSEL 5.0 program to exclude those with > 20 % missing data, non- polymorphic, distorted and unusual SNPs, a total of 3,308 SNPs remained and were used for association mapping (Bradbury et al., 2007). The physical positions of these markers in the soybean genome were determined using the whole genome assembly of Williams 82 soybean (Gmax 2.0).

With the 3308 SNPs genotype data, principle component analysis (PCA) was made by TASSEL 5.0 program (Bradbury et al., 2007). Kinship coefficient matrix was used to provide an estimate of additive genetic variance using TASSEL 5.0 program (Bradbury et al., 2007; Zang et al., 2010). The physical map of markers was conducted by R/qtl using plotMap () function. In the present study, we used models where a general linear model (GLM), a mixed linear model (MLM) using TASSEL 5.0 and GAPIT R package produced *p* values to populate Manhattan plots (Yu et al., 2006; Zang et al., 2010). The significance of associations between SNPs and traits was based on Bonferroni correction and False discovery rate (FDR) analyses using PROC MULTTEST in SAS.

Results

The phenotypic distributions of 90 $F_{4:5}$ RILs and the parents for sucrose content for this study were summarized in Figure 5.1 in three locations such as Columbia MO (Latitude 38.90, Longitude -92.27), Novelty MO (Latitude 40.01, Longitude -92.19) and South Korea (Latitude 36.07, Longitude 128.38). The distributions of sucrose content for the population in the three locations were continuous and there was a range of sucrose content from 5.1 % to 12.1 % in Columbia, from 2.8 % to 10.0 % in Novelty and from 3.5 % to 12.2 % in South Korea (Figure 5.1). Transgressive segregation was showed in three locations as RILs having less than 4% or more than 10 % sucrose concentration in the dry seed (Figure 5.1). According to normality test using SAS, the phenotypic distributions in the Novelty and South Korea locations showed a normal distribution where P-values for Shapiro-Wilk were 0.4926 and 0.2753, respectively. The RILs population including parental lines matured within 10 days in 2015 (data not shown).

A total of 3,308 SNPs were identified as polymorphic markers and used for the construction of a physical map and an association mapping dataset (Figure 5.2; Table 5.1). The physical map of polymorphic markers was constructed using r/qtl and consisted of 20 chromosomes that covered 949 Mbp (table 5.1). The number of markers in each chromosome were summarized in Table 5.1.

A genome-wide association study (GWAS) was performed with the general linear model (GLM) using the Tassel 5.0 program (Figure 5.3). The GLM results of the GWAS for sucrose content were presented in Table 5.2. A significant association with sucrose content was determined by Bonferroni correction and False Discovery Rate (FDR) and observed for 11 SNPs located in chromosome 6 for Columbia MO location (Table 5.2; Figure 5.3.a). Phenotypic

variation in Columbia ranged from 19.3 % to 40.6 %. Twelve SNPs were significantly associated with sucrose content in chromosomes 6, 8 and 20 in Novelty MO (Table 5.2; Figure 5.3.b). Phenotypic variation in Novelty ranged from 21.5 % to 41.0 %. The sucrose content in South Korea were associated with five SNPs in chromosomes 6, 8 and 20, according to FDR which is less than 0.01 (Table 5.2; Figure 5.3.c) with LOD score which were more than 4.0. Phenotypic variation in South Korea ranged from 19.8 % to 29.1 %. RS2 genotype was used as a phenotype and analyzed with association mapping, and significant results for SNPs were identified on chromosome 6 (Figure 5.3.d). Fifteen four SNPs were significantly associated using RS2 genotype as phenotype on chromosome 05 according to Bonferroni correction ($P < 0.05$) (Data not shown). With GLM result, the two SNPs on chromosome o6. S1_260656029 and S1_260656177 were shared in the three locations (Table 5.2). Sucrose content associated with two SNPs having 'A' nucleotide had a high sucrose content on chromosome 6 (Table 5.2). In addition, two markers on chromosome 20 overlapped in three locations (S1_905090046 and S1_905090055) (Table 5.2). RILs with SNPs having 'A' nucleotide on chromosome 20 accumulated a high level of sucrose content (Table 5.2).

A GWAS was performed with the mixed linear model (MLM) using GAPIT (Figure 5.4) which greatly reduced the false positive rates as shown in quantile-quantile (QQ) plots. A SNP marker on chromosome 06 (S1_260656029) for sucrose content had a significant association in Columbia and Novelty according to MLM analysis (Data not shown).

With sucrose content, SNPs (S1_260656029 and S1_260656177) on chromosome 06 were aligned with raffinose synthase 2 gene which was located 374 Kbp away (Table 5.3). Between S1_352076324 and S1_352209650 on chromosome 08, sixteen genes were identified.

However, among these genes, DnaJ 10-related was highly associated with SNPs on chromosome 08 (S1_352076324 and S1_352209650) (Table 5.3). The closest genes near to a SNP on chromosome 20 were Glyma.20g033800 and Glyma.20g026800, which was 784 Kbp and 858 Kbp away from S1_905090046, respectively.

Discussion

Sucrose content is a variable component of seed composition. More sucrose was produced when there were cooler temperatures during the pod filling period (Bilyeu and Wiebold, 2016; Kumar et al., 2010). In the present study, the RIL population was developed with similar major maturity genotypes to ensure the RILs would mature during a specific short time for MG III or IV. RILs and parental lines were matured within 10 days in 2015 fall during harvest time. By doing this, the environmental effect was reduced to accumulate the sucrose content during the pod filling period to better identify the candidate genes associated with sucrose content in soybean. In addition, soybean for three locations in the present study was grown for maturity group III and IV. Although fixing the maturity alleles and growing the right maturity group, the phenotypic distribution in Columbia MO was different from the one in Novelty MO and South Korea. Another environmental effect may be involved in the three locations.

The genetic basis of the reduced RFO trait derived from PI 200508 was reported. They found that a novel allele contained a deletion of three base pairs within the raffinose synthase 2 coding sequence (RS2, Glyma06g18890) resulting in a deletion in a highly conserved region of a tryptophan residue at position 331 (*rs2W331-*) (Dierking and Bilyeu, 2008). A mutation of this region in soybean associates with the increased sucrose level according to the pathway of biosynthesis of RFO. Ultra-low RFO lies containing the mutant *rs2W331-* allele with variant alleles of raffinose synthase 3 gene (*RS3*, Glyma05g08950, *rs3SNP6*) have the ultra-low RFO phenotype, characterized by RFO levels below approximately 1 % of the seed (Chapter 2). In the present study, parental lines SGUL has the *rs2W331-* and *rs3snp6* alleles, whereas parental line 534545 has the *RS2* and *rs3snp6* alleles and a high seed sucrose content. The RIL population

segregated for the *RS2* alleles and the result of GLM and MLM method in the present study clearly identified the *RS2* allele as having a major significant effect on sucrose content. In addition, we had used a direct *rs2w331-* genotyping assay to determine *rs2* allele status for this RIL population and used this result as a phenotype to analyze association mapping using TASSEL 5.0. This result matched the associated SNPs on chromosome 06 to the RS2 gene. In this population, the most important gene for sucrose content was the *rs2* allele.

Previous work was conducted with genetic mapping studies for sucrose content in soybean. Maughan et al (2000) concluded that seventeen marker loci, mapping to seven different genomic regions (Chromosome 05, 07, 08, 13, 15, 19 and 20), were significantly associated with sucrose variation (Maughan et al., 2000). In addition, there are three QTLs on chromosome 07, 08 and 20 which have a total of 23.2% of phenotypic variation (Wang et al., 2014). In the present study, transgressive segregation and normal distribution was noted for sucrose content in three locations, which suggested multiple genes were contributing to sucrose content in soybean. We identified significant results on chromosome 06, 08 and 20. To combine these results, there are other important candidate genes for sucrose content on chromosome 08 and 20 as well. In chromosome 08, sucrose content is associated with hilum color on soybean seed coat. With investigation of hilum color, lines with black hilum color had more sucrose content than other in this population (Data not shown). For the RIL population, black hilum color may be associated with sucrose content. A candidate gene on chromosome 20 may be important to accumulate sucrose content in soybean. Further research will be required for an analysis of candidate genes on chromosome 20.

Table

Table 5.1. Summary of physical map of markers

Chromosome	number of markers	End position (bp)	Pericentromere	
			Start (bp)	End (bp)
1	216	56,831,624	5,320,000	47,200,000
2	135	48,577,505	16,900,000	38,100,000
3	150	45,779,781	5,480,000	31,600,000
4	450	52,389,146	9,650,000	42,800,000
5	61	42,234,498	7,450,000	27,000,000
6	138	51,416,486	17,900,000	47,300,000
7	79	44,630,646	18,000,000	34,700,000
8	106	47,837,940	23,300,000	39,900,000
9	174	50,189,764	6,330,000	37,300,000
10	56	51,566,898	6,950,000	36,800,000
11	106	34,766,867	13,000,000	29,000,000
12	81	40,091,314	9,300,000	32,700,000
13	74	45,874,162	.	9,390,000
14	176	49,042,192	9,830,000	43,600,000
15	356	51,756,343	15,800,000	47,100,000
16	68	37,887,014	7,330,000	27,800,000
17	90	41,641,366	14,100,000	37,300,000
18	328	58,018,742	17,200,000	42,700,000
19	314	50,746,916	4,560,000	34,300,000
20	150	47,904,181	5,390,000	32,400,000
	Total = 3308	949,183,385		

Table 5.2 SNP loci associated with sucrose content in three locations and RS2W331- genotype according to GLM.

SNP ID	Chromosome	Position[a]	LOD[b]	Mean of sucrose content (%) associated with SNP allele					r2	Bonferroni	False Discovery Rate
				A (n[c])	T (n)	C (n)	G (n)	Het (n)			
Columbia											
S1_260116477	6	14303423	4.10		6.3 (19)	8.2 (45)		7.7 (10)	23.2	0.26	0.026
S1_260345934	6	14532880	4.11	8.2 (33)	6.7 (33)			8.1 (8)	20.7	0.257	0.026
S1_260345941	6	14532887	4.11			6.7 (33)	8.2 (33)	8.1 (18)	20.7	0.257	0.026
S1_260364485	6	14551431	8.83	8.4 (49)			6.3 (19)	6.6 (9)	40.6	<.0001	<.0001
S1_260463630	6	14650576	4.05	8.4 (30)			6.8 (37)	7.7 (22)	19.3	0.291	0.027
S1_260545218	6	14732164	4.18			8.2 (27)	6.6 (36)	8.1 (22)	19.6	0.218	0.026
S1_260656029	6	14842975	7.70	8.5 (30)			6.5 (39)	7.9 (17)	33.6	<.0001	<.0001
S1_260656177	6	14843123	4.56	8.4 (36)			6.5 (36)	7.9 (17)	21.4	0.091	0.018
S1_260714660	6	14901606	4.27	8.7 (27)			6.9 (42)	7.3 (15)	21	0.176	0.026
S1_261087191	6	15274137	5.74	6.7 (30)			8.6 (28)	7.7 (16)	28.6	0.006	0.002
S1_261514250	6	15701196	5.22	6.7 (37)		8.6 (33)		7.3 (16)	24.6	0.02	0.005
Novelty											
S1_260345934	6	14532880	3.87	7.6 (33)	5.7 (33)			6.9 (18)	23	0.447	0.044
S1_260345941	6	14532887	3.87			5.7 (33)	7.6 (33)	6.9 (18)	23	0.447	0.044
S1_260346117	6	14533063	3.83	5.7 (37)			7.5 (39)	6.6 (15)	21.5	0.483	0.044
S1_260364485	6	14551431	4.23	7.5 (49)			5.7 (19)	6.0 (9)	27	0.196	0.033
S1_260463630	6	14650576	4.26	7.9 (30)			5.9 (37)	6.6 (22)	23.7	0.183	0.033
S1_260656029	6	14842975	7.95	8.1 (30)			5.5 (39	6.8 (17)	41	<.0001	<.0001
S1_260656177	6	14843123	4.91	7.8 (36)			5.7 (36)	6.5 (17)	27	0.04	0.014
S1_261087191	6	15274137	3.75	5.9 (39)			7.9 (28)	6.3 (16)	23.7	0.582	0.049
S1_352209650	8	10349264	4.61	6.2 (37)			7.6 (37)	5.3 (12)	27.4	0.081	0.02
S1_905090046	20	3808942	3.83	7.5 (32)			6.1 (34)	5.1 (9)	26.2	0.486	0.044
S1_905090055	20	3808951	3.83	7.5 (32)		6.1 (34)		5.1 (9)	26.2	0.486	0.044
S1_908506765	20	7225661	5.10	6.0 (47)			8.2 (22)	5.2 (6)	32.9	0.026	0.013

South Korea											
S1_260656029	6	14842975	5.73	9.4 (30)			6.6 (39)	8.0 (17)	29.1	0.006	0.006
S1_260656177	6	14843123	3.98	9.0 (36)			6.7 (36)	7.8 (17)	20.1	0.347	0.087
S1_352076324	8	10215938	3.83		8.5 (53)	6.8 (25)		6.6 (11)	19.8	0.49	0.093
S1_905090046	20	3808942	4.04	8.8 (32)			7.2 (34)	6.2 (9)	23.8	0.3	0.087
S1_905090055	20	3808951	4.04	8.9 (32)		7.2 (34)		6.2 (9)	23.8	0.3	0.087

[a]Position in base pairs for the peak SNP according to soybean reference sequence of Williams 82; [b] LOD (logarithm of odds) indicates -Log_{10}P; [c]the number of RILs

Table 5.3. the candidate genes associated with sucrose content in three chromosomes.

SNP ID	Chr[a]	Position[b]	Gene[c]	Gm gene model[d]	Start (bp)	End (bp)	Distance from SNP (bp)
S1_260656029	6	14,842,975	Raffinose Synthase 2	Glyma.06g179200	15,217,419	15,223,877	374,444
S1_260656177	6	14,843,123	Raffinose Synthase 2	Glyma.06g179200	15,217,419	15,223,877	374,296
S1_352076324	8	10,215,938	DnaJ 10-related	Glyma.08g134400	10,299,402	10,307,179	91,241
S1_352209650	8	10,349,264	DnaJ 10-related	Glyma.08g134400	10,299,402	10,307,179	42,085
S1_905090046	20	3,808,942	Nitrate, Fromate, Iron Dehydrogenase	Glyma.20g033800	4,593,837	4,595,977	784,895
			Purple acid phosphatases superfamily protein	Glyma.20g026800	2,946,177	2,950,346	858,596

[a]Chromosome; [b]Position in base pairs for the peak SNP according to soybean reference sequence of Williams 82; [c]The gene located near to SNPs as reported previously; [d]Glycine max Wm82.a2. v1.

Figure

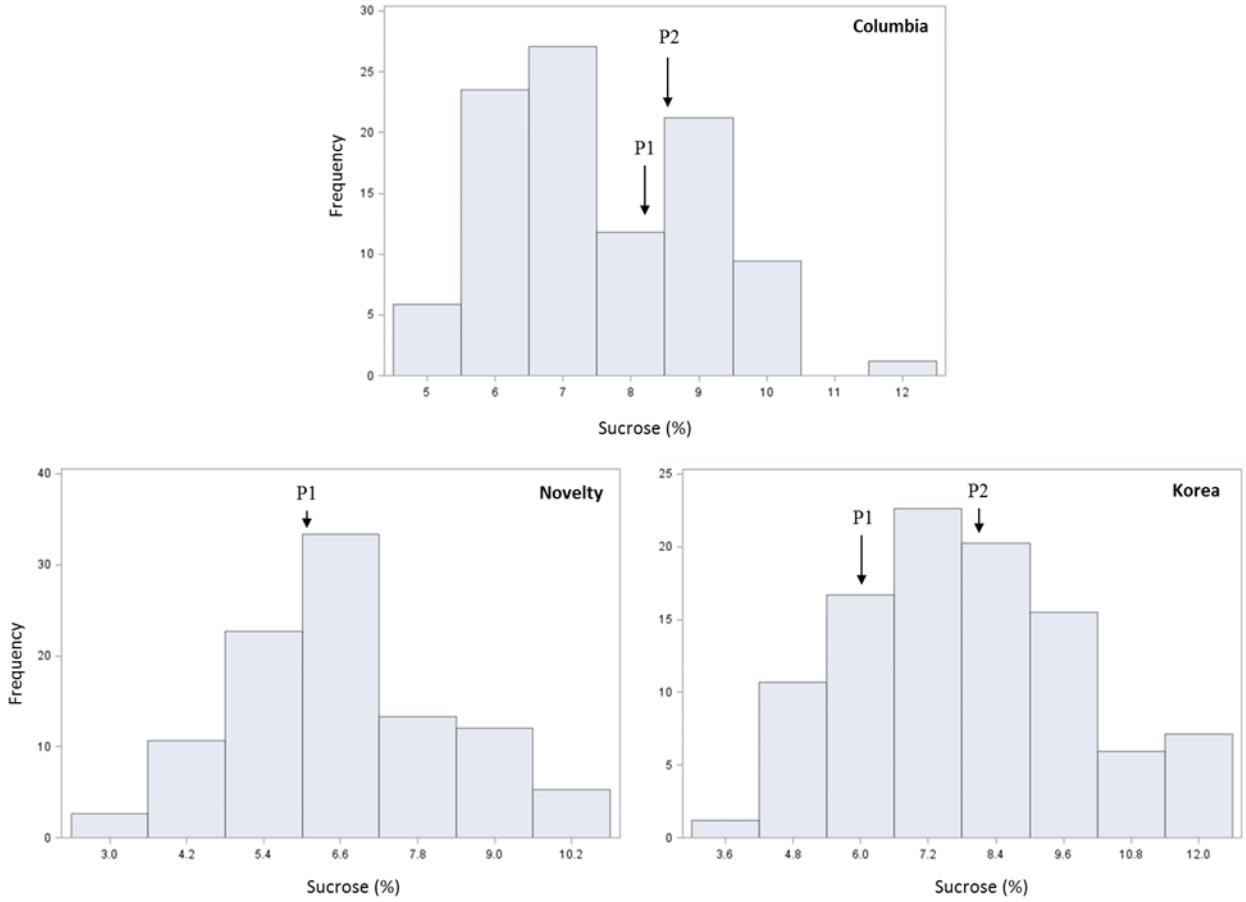


Figure 5.1. Frequency distribution of sucrose content in three locations such as Columbia, Novelty and South Korea. P1 represents SGUL and P2 represents 534545.

P2 in Novelty had a missing as poor germination.

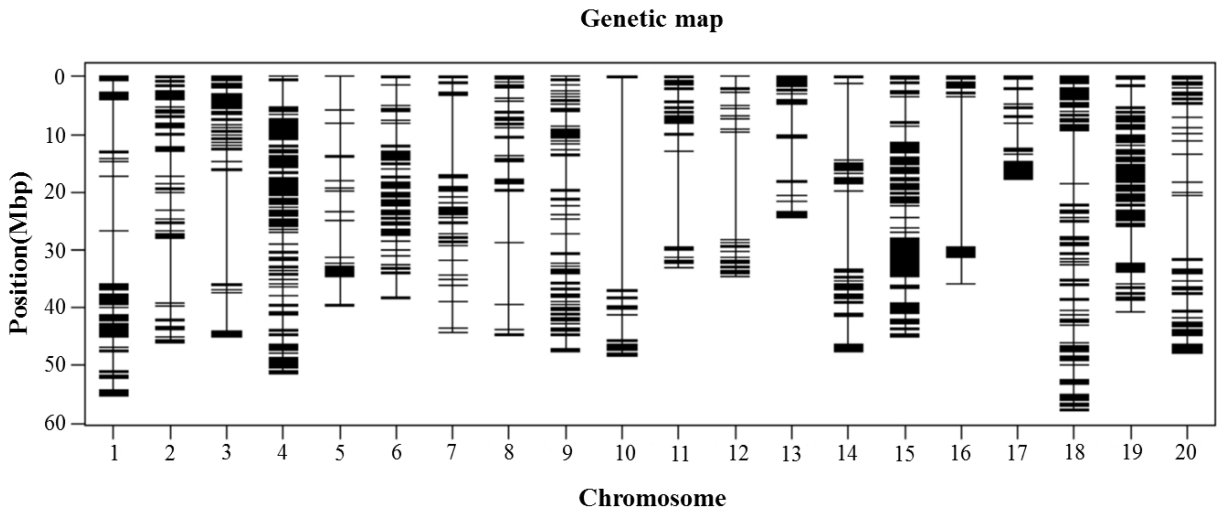


Figure 5.2 Physical map of markers based on polymorphism between 534545 and SGUL

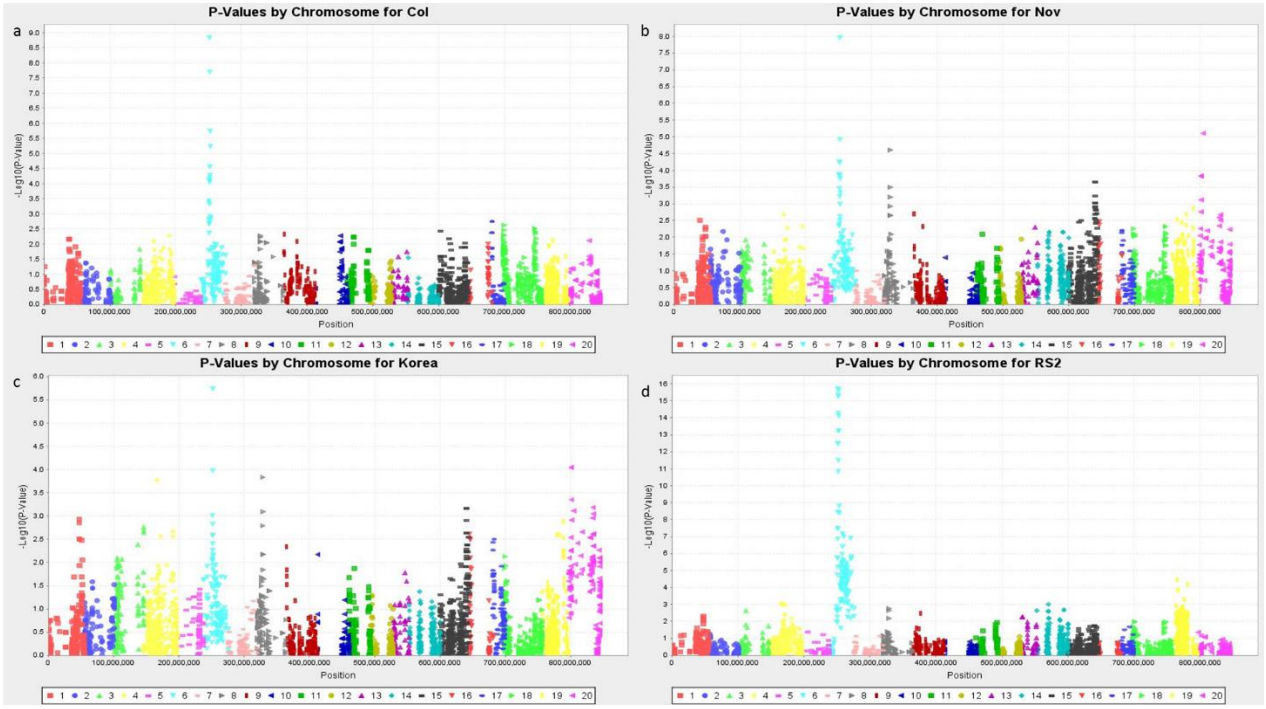


Figure 5.3 GWAS results of GLM method for sucrose content in three locations and rs2 genotype as phenotype using TASSEL 5.0. a. Columbia MO b. Novelty MO c. South Korea d. rs2 genotype as phenotype.

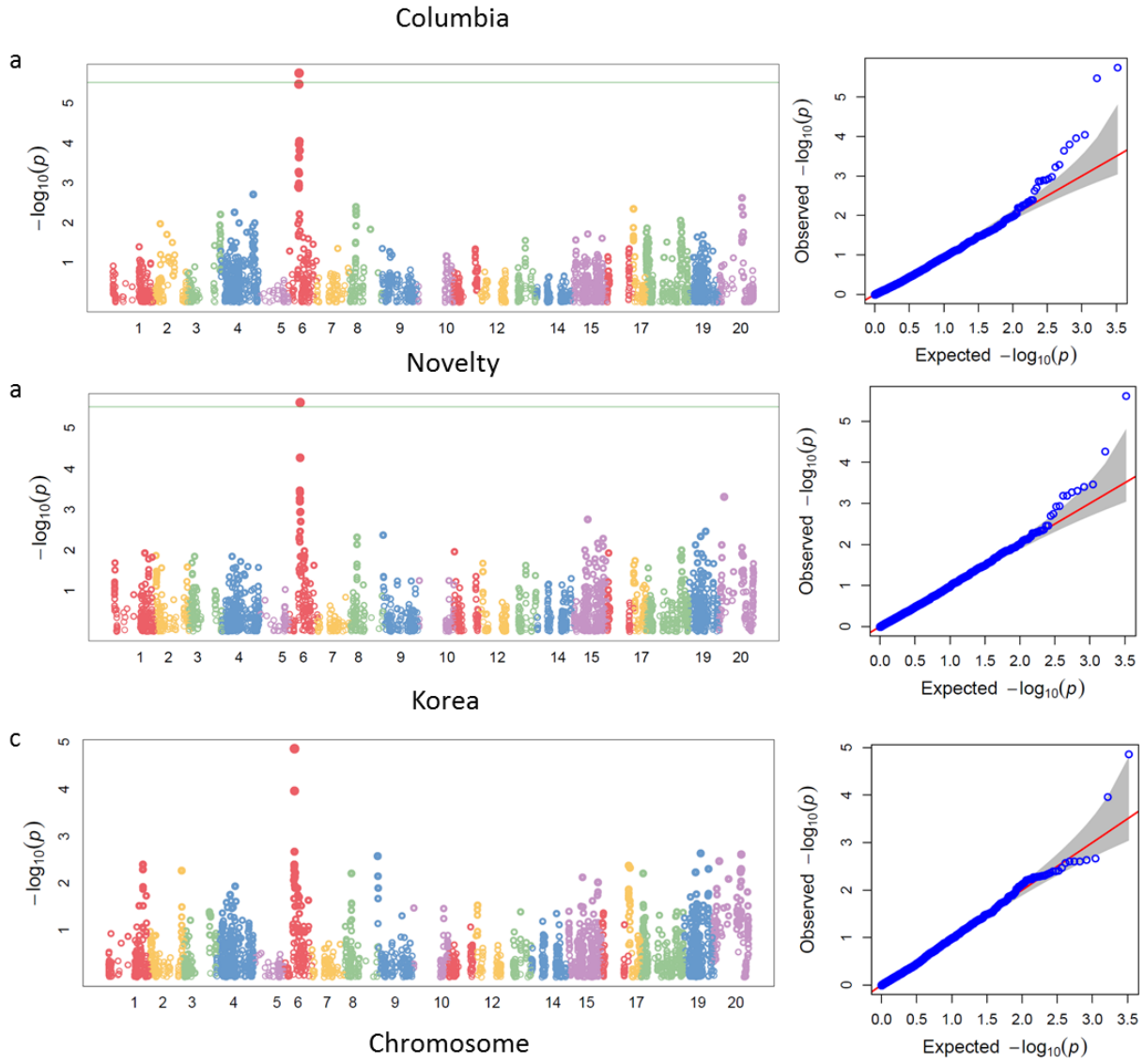


Figure 5.4. Manhattan plots and Quantile-quantile (QQ) plot of MLM method for sucrose content in three locations using GAPIT. a. Columbia b. Novelty c. South Korea. The horizontal green line indicates the genome-wide significance threshold (FDR < 0.05).

References

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.

Brown, S. (2006) Effect of Environment on Seed Composition of Tofu and Natto Soybean Cultivars. Master's thesis, University of Missouri, Columbia, MO, USA

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(19), pp.2633-2635.

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

Elshire RJ, Glaubitz JC, Sun Q, et al (2011) A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. PLoS ONE 6:e19379

Hymowitz, T. and Collins, F.I., (1974). Variability of Sugar Content in Seed of Glycine max (L.) Merrill and G. soja Sieb. and Zucc. Agronomy Journal, 66(2), pp.239-240.

Karr-Lilienthal L.K., Grieshop C.M., Spears J.K., Fahey Jr. G.C. (2005) Amino acid, carbohydrate, and fat composition of soybean meals prepared at 55 commercial U.S. soybean processing plants. J Agric Food Chem 53:2146-2150

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(1), 106-112.

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(5), pp.533-538.

Kumar, V. Rani, A.; Goyal, L.; Dixit, A. K.; Manjaya, J. G.; Dev, J.; Swamy, M. (2010) Sucrose and raffinose family oligosaccharides (RFOs) in soybean seeds as influenced by genotype and growing location. J. Agric. Food Chem., 58 (8), 5081−5085.

Maughan, P. J., Maroof, M. S., and Buss, G. R. (2000). Identification of quantitative trait loci controlling sucrose content in soybean (Glycine max). Molecular Breeding, 6(1), 105-111.

Obendorf, R.L., Horbowicz, M., Dickerman, A.M., Brenac, P. & Smith, M.E. (1998b). Soluble oligosaccharides and galactosyl cyclitols in maturing soybean seeds in planta and in vitro. Crop Science 38 (1), 78-84

Swarts K, Li H, Romero Navarro JA, et al (2014) Novel methods to optimize genotypic imputation for low-coverage, next-generation sequence data in crop plants. The Plant Genome, 7

Schillinger, J.A., Dierking, E.C. and Bilyeu, K.D., Schillinger Genetics, Inc. and The United States Of America As Represented By The Secretary Of Agriculture, (2013). Soybeans having high germination rates and ultra-low raffinose and stachyose content. U.S. Patent 8,471,107.

Teixeira, A.I., Ribeiro, L.F., Rezende, S.T., Barros, E.G. and Moreira, M.A., (2012). Development of a method to quantify sucrose in soybean grains. Food Chemistry, 130(4), pp.1134-1136.

Yu, J., Pressoir, G., Briggs, W.H., Bi, I.V., Yamasaki, M., Doebley, J.F., McMullen, M.D., Gaut, B.S., Nielsen, D.M., Holland, J.B. and Kresovich, S., (2006). A unified mixed-model method for association mapping that accounts for multiple levels of relatedness. Nature genetics, 38(2), pp.203-208.

Wang, Y., Chen, P. and Zhang, B., 2014. Quantitative trait loci analysis of soluble sugar contents in soybean. Plant Breeding, 133(4), pp.493-498.

Wilson RF (2004) Seed composition. In: Boerma HR, Specht JE (eds) Soybeans: improvement, production and uses. ASA, CSSA and SSSA, Madison, WI, pp 621-677

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

Zhang, Z., Ersoz, E., Lai, C.Q., Todhunter, R.J., Tiwari, H.K., Gore, M.A., Bradbury, P.J., Yu, J., Arnett, D.K., Ordovas, J.M. and Buckler, E.S., (2010). Mixed linear model approach adapted for genome-wide association studies. Nature genetics, 42(4), pp.355-360.

Zeng, A., Chen, P., Shi, A., Wang, D., Zhang, B., Orazaly, M., Florez-Palacios, L., Brye, K., Song, Q. and Cregan, P., (2014). Identification of quantitative trait loci for sucrose content in soybean seed. Crop Science,54(2), pp.554-564.