5-2012

Identification and Confirmation of SSR Markers and QTL for Seed Calcium Content and Hardness of Soybean

Moldir Orazaly
University of Arkansas, Fayetteville

IDENTIFICATION AND CONFIRMATION OF SSR MARKERS AND QTL FOR SEED CALCIUM CONTENT AND HARDNESS OF SOYBEAN

IDENTIFICATION AND CONFIRMATION OF SSR MARKERS AND QTL FOR SEED CALCIUM CONTENT AND HARDNESS OF SOYBEAN

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy in Cell and Molecular Biology

By

Moldir K. Orazaly
Al-Farabi Kazakh National University
Bachelor of Science in Biology, 2003
Al-Farabi Kazakh National University
Master of Science in Biology, 2005

May 2012
University of Arkansas

ABSTRACT

The health benefits and the nutritional values of soyfood products have expanded the demand for food-grade soybean varieties. For whole soyfood products such as natto, small seeded varieties with proper texture are essential. Seed calcium content and hardness are main factors in determining the seed texture. Utilizing molecular markers associated with quantitative trait loci (QTL) for marker assisted selection (MAS) has been an efficient technique for breeders. To find molecular markers associated with a trait is an initial step. The confirmation of QTL in different environments and in different genetic backgrounds is important for MAS. The objectives of this research were to identify SSR markers and QTL associated with seed calcium content and hardness, to confirm previously reported QTL in different genetic backgrounds using mapping populations and advanced breeding lines, and to identify correlation between seed calcium content and hardness. RILs derived from KS4303sp (low calcium/soft-seeded) x PI 407818 B (high calcium/hard-seeded) and PI 407818 B (high calcium) x PI 408052 C (low calcium) were analyzed to identify SSR markers and QTL for seed calcium content and hardness. These RIL populations were grown in a randomized complete block design (RCBD) with two replications, at Fayetteville, AR and Keiser, AR in 2008 and 2009. Interval mapping for the SSR markers revealed a new putative QTL *Ca5* for seed calcium content. This QTL is linked to the markers Sat_290 and Satt115 on chromosome 18. In addition, regions on chromosomes 1, 2, 5, 7, 8, 18, and 20 contained environmentally stable markers significantly associated with seed calcium content.

RIL population derived from KS4303sp x PI 407818 B was used to identify SSR markers and QTL associated with seed hardness. Interval mapping for the SSR markers revealed a new putative QTL associated with seed hardness. This QTL is linked to the markers Satt547 and

Satt414 on chromosome 16. Single marker analysis also revealed 12 markers associated with seed hardness on 7 different chromosomes of soybean.

One hundred sixty four advanced breeding lines derived from six different crosses with contrasting seed calcium content and hardness and the population derived from KS4303sp x PI 407818 B were used to identify correlation between these two traits. Most of the populations showed positive correlation between seed calcium content and hardness. One hundred sixty four advanced breeding lines from six crosses also were used to confirm previously reported and newly identified QTL for seed calcium content and hardness. Previously reported calcium QTL (*Ca1*, *Ca2*, and *Ca4*) and newly identified QTL (*Ca5*) were confirmed in these advanced lines. In addition, markers Satt267 and Sat_345 on chromosome 1, Sat_288 on chromosome 7, Sat_228, Satt341, Sat_392 on chromosome 8, Satt547 on chromosome 16, and Satt002 on chromosome 17 were most consistent markers associated with calcium content across populations and environments. Previously reported hardness QTL (*Ha1* and *Ha2*) and newly identified QTL (*Ha3*) were confirmed in these advanced lines. Moreover, markers Satt267 on chromosome 1, Satt680 on chromosome 7, Satt341 on chromosome 8, and Sct_010 on chromosome 19 were stable across populations and environments for seed hardness. Most consistent markers across populations and environments are the best candidates for MAS. Findings of this research will facilitate MAS for seed calcium content and hardness in soybean breeding programs for food grade soybeans.

This dissertation is approved for recommendation
to the Graduate Council.

Dissertation Director:

_____

Dr. Pengyin Chen

Dissertation Committee:

_____

Dr. Navam Hettiarachchy

_____

Dr. Esten Mason

_____

Dr. Ines Pinto

DISSERTATION DUPLICATON RELEASE

I hereby authorize the University of Arkansas Libraries to duplicate this thesis when needed for research and/or scholarship.

Agreed _____

Moldir K. Orazaly

Refused _____

Moldir K. Orazaly

ACKNOWLEDGEMENTS

I would like to thank my academic advisor Dr. Chen for his guidance, support, and encouragement and giving me the opportunity to be part of the soybean breeding program at the University of Arkansas. During my journey at the graduate school he has been supportive as an academic advisor and understanding as a friend who is ready to give advices in science as well as in daily life. For all that, I cannot thank him enough.

I am very thankful for my committee members Dr. Pinto, Dr. Hettiarachchy, and Dr. Mason for serving on my committee and for their suggestions, support, and advice.

I am thankful to Bo Zhang who is an excellent supervisor who has been there to help and to advise.

I would like to thank all of the current and past members of soybean breeding program at the University of Arkansas. Special thanks to Caroline Grey, Leandro Mazzoni, Tetsuaki Ishibashi, Julietta Alaotti, Ehsan Shakiba, Yoko Yoshikawa, John Carlin, Tina Hart, Innan Cervantes, Luciano Jaureguy, Ailan Zhang, and Liliana Florez who have been a tremendous help during planting, harvesting, and data analysis. All members of the soybean breeding team and new additions make working environment enjoyable.

I am very thankful to employees of Arkansas Agricultural Research and Extension Centers at Fayetteville and Keiser, AR for their help in field experiments and maintenance.

I thank the Department of Crop, Soil, and Environmental Sciences and Cell and Molecular Biology Program at the University of Arkansas where I received generous help whenever I needed.

I owe my deepest gratitude to the Presidential Scholarship “Bolashak” from Kazakhstan that supported me financially and gave me the opportunity to study abroad.

I would like to thank all the employees of International Students and Scholars Office at the University of Arkansas especially Catherine Cunningham who was always there to help.

I am extremely thankful to my friends who have been supporting, loving, and believing all these years while I was away from my family.

Most importantly, I would like to thank my family who believed in me and showed endless support during my PhD program.

DEDICATION

The doctoral dissertation is dedicated to my father Koyanbay Orazaly and mother Kanziya Elaman, whose endless love, support, and encouragement made me who I am today, and to the memory of my brother Nurlan Koyanbayuly.

TABLE OF CONTENTS

Chapter 5

LIST OF TABLES

Chapter 2

Chapter 3

Table 4. Mean effect of SSR marker alleles on seed hardness (in newtons) in 120 RIL from the cross KS4303sp x PI 407818 B in 2008 and 2009.

Chapter 4

Table 1. Pair-wise analysis of Pearson's correlation coefficient for average calcium content and seed hardness in a 120 RIL population derived from the cross KS4303sp x PI 407818 B evaluated at Fayetteville, AR (FAY) and Keiser, AR (KEI) in 2008 and 2009.

Table 2. Pair-wise analysis of Pearson's correlation coefficient for average seed calcium content and hardness in 164 breeding lines derived from six crosses evaluated at Fayetteville, AR (FAY) in 2007 and Fayetteville, AR and Keiser, AR (KEI) in 2008.

Table 3. Average seed calcium content of parental lines and 164 breeding lines derived from six crosses evaluated in two locations (Fayetteville, AR and Keiser, AR) and over years (2007 and 2008).

Table 4. Average seed hardness of parental lines and 164 breeding lines derived from six crosses evaluated in two locations (Fayetteville, AR) in 2007 and (Fayetteville, AR and Keiser, AR) in 2008.

Table 5. Summary of polymorphic SSR markers used in screening 164 breeding lines derived from six crosses (Hutcheson x Camp, SS516 x Camp, SS-516 x V96-4486, Hutcheson x V96-4486, MFS-591 x Camp, and MFS-591 x V96-4486) for association analysis.

Table 6. Single marker analysis of variance for seed calcium content of 164 breeding lines derived from six crosses evaluated in Fayetteville, AR (FAY) and Kibler, AR (KIB) in 2007 and Fayetteville, AR and Keiser, AR (KEI) in 2008.

Table 7. Single marker analysis of variance for seed hardness of 164 breeding lines derived from six crosses evaluated in Fayetteville, AR (FAY) in 2007 and Fayetteville, AR and Keiser, AR (KEI) in 2008.

LIST OF FIGURES

Chapter 2

Chapter 3

Chapter 4

Chapter I

Introduction and Literature Review

Soybean production

Soybean [(*Glycine Max* (L.) Merr.] is one of the most valuable crops in the world, not only as an oil seed crop and feed for livestock, but also as a good source of protein for the human diet and as a bio-fuel feedstock (Smith and Huyser, 1987). It's been grown in Asia since ancient times and has since been cultivated around the world. In 2010, 258.4 million metric tons of soybeans were produced worldwide with 35 % of the production belonging to the US followed by Brazil, Argentina, and China. As a good source of oil, soybean accounted for 58% of the oil production in the world in 2010 (Soystats, 2010). In 2010, soybean export of the United States represented 44% of world soybean trade. The largest customers were China followed by Mexico, and Japan (Soystats, 2010). In the US, the leading soybean producing states are Iowa and Illinois with 13.51 and 12.68 million metric tons soybean production on 3.966 and 3.683 thousand hectare area, respectively (Soystats, 2010). In Arkansas, 1.3 thousand hectare of soybeans was planted in 2010 and producing 3.0 million metric tons of soybeans with 2.35 metric tons per hectare yield.

Soybean seed composition

Dry soybean seed contains approximately 40% protein, 20% oil, 35% carbohydrates, and 5% ash. The World Health Organization (WHO) indicated that soy protein has the minimum daily requirement of all amino acids, including amino acids limited in other food sources such as methionine and cystine (World Health Organization, 2008). On average, 100 g of soybean oil contains 16 g of saturated fat, 23 g of mono unsaturated, and 58 g of polyunsaturated fat. The major soybean unsaturated fatty acids are linolenic acid and oleic acid. Saturated fatty acids in soybean seed are stearic and palmitic acid. The major soluble carbohydrates of soybean are composed of 5-7% sucrose, 0.5-1.3% raffinose, and 2-4% stachyose (Guillon and Champ, 2002).

Mineral content of soybean, determined as ash, includes iron, calcium, zinc, phosphorus, magnesium, sulfur, chloride, and sodium, all in the trace amount (Lui, 1997).

Products and uses

A high content of protein and oil in the seed makes soybean an important source for human and livestock nutrition in North and South America, Europe, and Asia (Liu, 1997). Various soybean food products are used for human consumption, such as fermented soyfood, soy beverage, flour, and whole-bean products. Textural vegetable protein from soybean is used as stimulated meat, fruit, and nut products (Chomachalow, 1993). After processing, seeds of soybean are used as a valuable, protein-rich feed supplement for livestock (Fageria et al., 1997). Current market provides consumers with a variety of soy foods such as traditional soy foods, soy oil products, soy protein products, modern soy foods, soy enriched foods, and soy dietary supplements. Soy paste, soy souce, tempeh and natto are considered the major fermented soyfoods. The main non fermented soyfood include soymilk, tofu, and soy sprouts. Vegetable soybeans have a higher amount of fiber, iron, calcium, vitamin C and protein than any other types of beans (Liu, 1996). Soybean oil is used in many ways, such as in various pharmaceuticals and medicines, and in manufacturing disinfectants, printing inks, lubricant, wax, and soaps.

Soyfood

Soyfood is divided into types such as products derived from small or large seed and fermented or non-fermented products. Small seeds are desirable for natto (fermented soybeans), soy sauce, and soy sprouts which require <12g/100 seeds (Geatler et al., 2000). Natto and soy sprouts are produced using not only small seeds but also using lines with high fermentable sugars. Tofu (soybean curd), edamame (green vegetable soybeans), and soymilk are produced

using large seeds (>20g/ 100 seeds). Seeds with high sucrose and high protein content are desirable.

Vegetable soybean

Vegetable soybean is called “edamame” in Japan and “mao dou” in China and it is a very common food in Asian countries. Lately, it has become popular in the U.S. as well. Due to the increasing demand, there is a new “Vegetable soybean industry” initiated in Arkansas which will start to produce vegetable soybeans in large scales in 2012. Vegetable soybean is harvested at the R6 (full pod) physiological stage (Fehr et al., 1971) unlike regular commodity soybean which is harvested at the R8 (full maturity) stage. It is a product with a sweet and nutty flavor. It is a very nutritious food used in dishes such as salad, soup, and stews. It is used as a good source of protein (Miles et al., 2000), calcium, vitamin A, vitamin C (ascorbic acid), vitamin E (tocopherol), and dietary fiber (Johnson, 2000), and it is low in trypsin inhibitors, oligosaccharides, and phytate (Liu, 1996). In the U.S., it is sold in different forms such as fresh or frozen, in pods or shelled. Vegetable soybean is served plain, salted, or spiced after boiling for 10 min. In order to meet the requirements of vegetable soybean, lines with light green pods, large seeds with soft and nutty texture, uniform maturity, and sweet mild flavor are used (Miles et al., 2000).

Soymilk

Soymilk is originated from China and used in Asian countries for centuries. It was also adopted in other countries after eliminating beany flavor. Soymilk is made by grinding dry whole soybean seed into a powder with particle size < 420 microns. It can also be made by soaking whole beans for several hours and then grinding it wet followed by filtering then cooking or heating. The seed soaking process reduces phytate levels by leaking into the water. Cooking

soybeans improves the digestibility of seeds, inactivates trypsin inhibitors, and denaturates storage proteins. Heating also lowers viscosity, which improves obtaining proteins and solids. Advantages of soymilk include no cholesterol, no lactose, easy digestibility and high nutritional value. Soymilk contains 90% water, 3% protein, 1.5% lipids, and 1.5% carbohydrates. In contrast to dairy milk, it contains high protein, more unsaturated fatty acids, more iron with less fat, les carbohydrates, and low calories. (Lucas et al., 1989). Soymilk comes in plain, flavored, condensed, and powdered varieties. Soybean milk production is simple, yet there are important criterions to meet such as less beany flavor, smooth mouth feel, high yield and good taste rating in addition to using large seeded soybean lines with high protein and high yield.

Tofu

Tofu is made by precipitation of the proteins with a salt ($CaSO_4$ or $MgCl_2$) or acid (glucono-δ-lactone) from soymilk. Soymilk is heated to cause protein dissociation and adding coagulants gives firmness. It is creamy white colored, smooth textured, has a mild taste, and is a highly hydrated gelatinous product (Wang et al., 1983). Due to its fine taste it is one of the well accepted soyfood in North America. Its role in breast cancer prevention may be another reason of wide consumption of the product (Shu et al., 2001). It contains 88% moisture, 6% protein, 3.5% lipids, and 2% carbohydrates (Min et al., 2005). Tofu is low in calories, has a relatively significant amount of iron and small amount of fat. Soymilk that is used to make tofu should be light colored. Usually soybean lines with large seeds and light colors with high protein and high sucrose content are used to make soymilk and tofu. Selection of soybean lines for tofu production is based on traits such as protein, sugar, oil content, seed size, and seed color. Water absorption ratio is also important. According to Poysa et al., (2006) yield of soymilk and tofu were positively correlated with seed protein content and stachyose; while negatively correlated

with seed oil and sucrose. The texture of tofu is very important and it is based on two major storage protein components: 7S globulin (β-conglycinin) and 11S globulin (glycinin), which play a significant role in the gelation process of soymilk. Based on the ratio of these protein components and coagulants, pore size of tofu will vary (Kohyama et al., 1995). The 7S/11S ratio has a major effect on the physical properties of tofu gel. It also affects firmness of tofu and breaking stress values. Information about traits such as chemical components of soybean, gelation rate, and gel properties in tofu processing is helpful to breeders in manipulating soybean genetics through breeding.

Natto

Natto is a fermented soybean product mostly consumed in Asian countries and considered as traditional food in Japan as a breakfast meal. It is fermented with a bacterium *Bacillus subtilis* which gives it a unique taste and aroma that it is not yet favored by many places other than countries of origin. Even though natto is not consumed all around the world, large amount of natto soybeans are produced in the US and exported to Asian countries. For natto production, soybean lines with small round shaped seeds with yellow hilum color are required. Before making natto, seeds are screened for size, color and cleaned from contaminations and broken seeds. Selected seeds are washed and then soaked for up to 20 hours at 12^{o}C. During this time seeds will absorb the water and increase in size. Soaked seeds will be pressure cooked at 1.2 kg/cm^2 and 120^{o}C for 20 min. This step not only softens the seeds, it also eliminates undesirable bacteria. Due to high temperature tolerance of *Bacillus subtilis*, pressure cooked soybeans will be inoculated while they are hot. During this process seeds will obtain stickiness. It is followed by packaging and fermentation. The fermentation process requires 40^{o}C and high humidity that is lowered gradually throughout 24 hours and then cooled. Natto is a great source of protein,

isoflavones, vitamins, and minerals. It contains 40 g protein, 25 g of lipid, and 30 g of carbohydrates on average in 100g of natto (Hosoi and Kiuchi, 2003). Using natto helps to prevent heart attack and cancer (Hosoi and Kiuchi, 2003). It also has high amount of menaquinone-7 which is known for being helpful for bone stiffness and for maintaining bone mineral density of human. Studies show that natto is one of the candidates that may account for difference in hip fracture and has positive association with bone mineral density (Katsuyama et al., 2002 and Ikeda et al., 2006).

There are seed traits and attributes that are important for natto production. Water absorption and cooked seed hardness are one of the main characteristics of natto. Since high water absorption is required in order to have soft textured natto, it is better to use lines with low oil content since high oil prevents water absorption. Natto hardness is negatively correlated with sugar content (Greater and Fehr, 2000). Unlike sucrose, stachyose is desirable and it is associated with small seed size (Taira, 1990). High calcium content causes harder seeds while low calcium seeds are softer (Chen et al., 2001). It can be due to the structural role of calcium on the cell membrane (White and Broadley, 2003). Seeds with high calcium content cause "stone seeds" that do not absorb water, and are therefore undesirable for the soyfood market because it is considered poor quality and costs extra to remove them (Zhang et al., 2008).

Seed hardness

Hardness, brittleness, and gumminess are the important characteristics that affect food quality such as tofu and natto. Such quality characteristics of soyfood are associated with water intake of dry seed and hardness of cooked seed. It is also important for the germination process in crop production that the seed coat have high permeability. Seeds with low imbibition can stay dormant for a long time which is not economically desirable. There are several factors that cause

hard seededness or “stone seeds”. Hardseededness of soybean is caused by impermeability of seed coat. In such case, hard seeds do not absorb water for up to eight days (Potts, 1978). Hard seededness of crops has both biological and economic impacts. There are both negative and positive impacts of this trait (Qutob et al., 2008). Beneficial effects of hardseededness are preserving seed stock for many years (Tyler, 1997), maintaining low seed moisture with high viability, and resistance to seed coat pathogens. Even though the effects listed above seems evolutional, hardseededness causes issues such as lack of uniform hydration where it is very important for processing soymilk, tofu, and miso (Shao et al., 2007). It causes additional cost for thermal, chemical, or physical treatments (Argel and Patron, 1999) because seeds with slow germination may require scarification. In addition, seed hardness refers to harder seeds with harder texture although they absorb water.

There are several reasons that explain the cause of seed hardness where seeds absorb water, yet not soft such as chemical composition of seed (Shao et al., 2007) and structural and permeability properties of seed coat (Qutob et al., 2008). Impermeability is associated with thick cuticles, thickness of testa, and calcium content (Werker et al., 1979). A close relationship has been reported between calcium content and the hardness of cooked bean (Saio et al., 1973). Among other traits, phytate is also correlated with seed hardness (Mullin and Xu 2001). Insoluble carbohydrates, hemicelluloses, cellulose, and pectin are major components of soybean seed coat. According to Mullin et al., (2001) low pectin content and high hemicellulose reduces hydrophilicity of seed coat which causes stone seeds. Protein content is low in the seed coat which is mainly involved in defense against pathogens whereas embryo contains all the storage proteins (Dhaubhadel et al., 2005). Depending on seed coat composition, soybean seeds may have dull or shiny and glossy luster. It was indicated that luster types are based on the presence

and density of deposits on the seed surface which are from endocarp tissues of seed pod (Yaklich et al., 1986). They are basically hydrophobic protein from soybean (HPS) derived from prolamin family and similar to lipid transfer proteins (LTP). HPS triggers the attachment of endocarp to the seed and affects seed surface structure such as luster and other physical properties (Gijzen et al., 2003).

The positive association of seed impermeability and seed pigmentation loci (Keim et al., 1990) indicates possible genetic linkage between these traits (Qutob et al., 2008). Seed pigmentation genetics involves genes associated with anthocyanin and proanthocyanidin in the seed coat and hilum (Todd and Vodkin, 1993). Loci *T*, *I*, and *R* in chromosomes 6, 8, and 9 respectively are known for pubescence color, hilum color, and seed coat pigmentation. Combination of recessive *i* and *t* alleles cause seed coat with cracks (Zabala and Vodkin, 2003) which may explain the correlation of seed coat color and seed hardness due to impermeability. Studies show positive correlation between seed impermeability and pigmentation loci (Keim et al., 1990) and seed hardness (Hou and Thseng, 1991). Qutob et al., (2008) suggested a possible pleiotropic effect between loci of pigmentation and permeability through their activity towards polyphenolic substrates that become impregnated in the cell wall.

Mullin et al., (2001) indicated the relationship between seed coat structure and hardness. When seed coat is mentioned with hardness, it is about water intake through the seed coat which is important as an initial step in cooked seed hardness. The role of the cuticle is significant in water movement. It consists of cutin polymer with waxes, terpens, proteins, fatty acids, and phenolics. Ma et al., (2004) compared six different varieties of soybean and found that small cracks in the cuticle covering palisade layer of soybean seed coat can explain permeability while varieties with no cracks were impermeable. Those microscopic cracks are not visible to the

naked eye and they are in average 1-5 μm wide and 20-200 μm long. Hard seeded lines had a strong cuticle and were resistant to breakage. In addition to seed coat structure, seed size was also found to be associated with permeability. Smaller seeds were less permeable and had thicker seed coat than large seeded lines (Calero et al., 1981). Even though much research has been done to understand seed hardness, our knowledge of seed hardness genetics is still lacking, which is one of the major concerns in soyfood production.

Seed calcium content

Calcium is an important nutrient for humans, animals, and plants due to its structural and signaling role. Low calcium intake can cause osteoporosis and hypertension in humans and poor biotic and abiotic stress tolerance and low yield in plants. Its role in plants is vast including roles as a nutrient, signaling element, and dietary source for humans and animals.

Recommended daily intake of calcium is 800-1300 mg for adults, yet daily human intake does not meet the recommended amount, especially in the developing countries. As a fundamental building block of bone, it is very important to obtain optimal calcium content during the stage of bone mass increase and also throughout life. Women with low calcium intake in childhood and adolescence have higher risk of fractures later in life (Kalkwarf et al., 2003). In order to meet the daily recommended calcium intake level, it is important to find cheap and easily available calcium sources. Dairy products are known calcium sources, but because of dietary preferences or dairy product intolerance, there could be limitations for some people and for those plant products can be another good source of calcium. Since micronutrient malnutrition has global effect (Cakmak, 2002), it is important to find good sources of plants with high nutrition. Legumes are one of the highest nutrient sources of protein and minerals. Vegetable calcium is more available to the human body due to its protein content because legume proteins

have lower amount of cystein and methionin compared to animal proteins (Young et al., 1994). The food and Drug Administration approved that foods that contain at least 6.25 g of soy protein per serving reduce the risk of cardiovascular diseases. Depending on food processing and cooking methods calcium content of soyfood can vary. For example, lactonic soybean curd contains 12.55mg/100g calcium while diced fried soybean curd contains 760mg/100g (Ma et al., 2005).

One major consideration about calcium content is its bioavailability. Even seeds with high calcium content may provide only a little amount of it due to the presence of "anti-nutrients" such as oxalate and phytate. Dayold et al., (2010) indicated that when phytate and/or oxalate content are high, calcium absorbability is low. Unlike experiments performed on rats, calcium bioavailability of soybean was highly influenced by phytate content due to the fact that humans lack intestinal phytase activity (Mason et al., 1993). Heaney et al., (1991) compared calcium bioavailability in dairy milk to soybeans with high and low phytate content. Low phytate soybeans had higher calcium availability than high phytate soybean. Calcium availability of dairy milk was lower than low phytate soybean but higher than high phytate soybean lines. It is also important to take into consideration that phytate may degrade during processing. According to Harland and Oberleas (1986), soaking and phytate loss have a positive correlation. Calcium content in soybean seed may depend not only on phytate content but also other minerals. Minarik and Shive (2005) indicated the influence of boron concentration on calcium; excess or deficiency of boron caused lower calcium level. Aluminum is another element that can reduce calcium by changing the membrane permeability (Rengel, 1992) and altering calcium homeostasis.

Calcium content in soybean seed is positively correlated with seed hardness yet it is not consistent due to environmental effects (Chen et al., 2011). It can be explained with its role in

cell membrane where most of the calcium is bound or sequestered in different organelles. Calcium is an abundant element in most environments yet it is immobile in plants except for in certain cases, such as during calcium deficiency in tissues, but can still only be re-mobilized in a very small amount (Keiser and Mullen, 1993). Since calcium is very important in cell physiology and structure, seeds can take more calcium than is needed and it can be stored in crystal form until the embryo needs it (Ilarslan et al., 2001). The cell wall provides a major sink for calcium and its crystals in plants (Kinzel, 1989). It may be possible that the high calcium content of the seed and its correlation with seed hardness can be explained by the crystal cell layer in the seed coat that is known for its role in embryo protection. Calcium acts as a transducer during cold or salt stress (Minorsky, 1985). Its role in strengthening cell wall and defense in pathogen attack has been suggested (White and Bradley, 2003). Aforementioned quality of calcium in seeds can be explained as an evolutional defense mechanism similar to the evolutional role of seed hardness especially in *Glycine soja* where calcium content (Raboy et al., 1984) and seed hardness are higher than in *Glycine max*. Another similarity between calcium content and seed hardness is their role in germination influencing longer shelf life (Doyd et al., 2010). Even though high calcium content sounds like a significant evolutional trait, it may negatively affect seed quality for whole seed soyfood products such as natto. Stone seeds, which are caused by high calcium level, are considered a major problem in the food industry. Its inconsistent correlation with seed hardness is explained by environmental effect (Chen et al., 1993) yet requires more research. Seeds with non-uniformed water absorption affect fermented soy food quality and removing stone seeds imposes additional cost in processing (Mullin and Xu, 2001). Practically, varieties with desirable calcium content and hardness are needed for soyfood production. Having control over calcium of soybean seeds by finding QTLs and using marker

assisted selection for this trait will improve breeding efficiency. There are a number of gene families that encode proteins involved in metal transportation in plants, but studies in the movement of minerals to seed by QTL analysis are limited to common beans (Sankaran et al., 2009).

Molecular markers in breeding

Genetic engineering and use of molecular technique gives an advantage for plant breeding to produce crop varieties with the desirable traits. Molecular markers such as random amplified polymorphic DNA (RAPD), restriction fragment length polymorphism (RFLP), amplified fragment length polymorphic DNA (AFLP), single nucleotide polymorphism (SNP), and simple sequence repeat (SSR) have been used for many crops to identify QTLs for economical and agronomical important traits such as yield, protein, oil, sugar, disease resistance, and others.

RAPD is based on amplification and visualization of random sections of DNA. RFLP is based on digestion of genomic DNA into fragments and determination of fragment size. AFLP is based on Polymerase Chain Reaction (PCR) amplification of restriction fragments generated by restriction enzymes and oligonucleotide adapters of few nucleotide bases. SNP is based on the differences in a single site in DNA (Sharp et al., 2001).

SSR markers or microsatellites are based on polymorphism of repeat units in a defined region of a genome. They usually consist of di, tri, and tetra base pair repeat units. In plants, the most common repeat unit is AT followed by AG, and TC. A nucleotide sequence flanking the repeat is used to design primers to amplify repeat units. Depending on the repeat number of the sequence, different sizes of DNA are amplified by PCR. Thus, differences of alleles between genotype can be visualized. Homozygous line will have the same DNA fragment from both

alleles, while if a line is heterozygous, it will generate DNA fragments of different sizes. Advantages of using SSR markers are their ubiquity, reproducibility, codominance, multi-allelism, high level of polymorphism, simple technique, low template DNA requirement, cost effectiveness, and high numbers of SSR primers that are publicly available (Edwards et al., 1996).

Quantitative trait loci (QTL)

Studying the domestication-related traits has shown that QTLs are not randomly distributed in the genome, they mostly are clustered in specific regions of the genome. Most of the economically important traits are quantitative in nature which means they are controlled by many genes and each one of them has small effect. Such traits are also affected by the environment unlike qualitative traits, which makes it harder to find those genes controlling a quantitative trait. The contribution of each quantitative locus at a phenotypic level can be expressed in an increased or decreased level of a trait. A quantitative trait has continuous variation with normal distribution. Since genetic and environmental factors contribute in an increasing and decreasing manner, looking at phenotypic distribution alone is not enough to identify QTLs. Analyzing population distribution, mean values, and genotypic, phenotypic, and environmental variances are necessary. This process is called QTL mapping. Finding polymorphic markers in a population and identifying the markers that are significantly associated with the trait can lead to QTL identification. A chromosomal region that is linked to the marker and affects the trait of interest is called a QTL. Finding QTL should be followed by confirmation and validation in different genetic backgrounds and environments before offering to marker assisted selection (MAS).

MAS is based on using DNA markers that are linked to a certain trait and its target loci. It can be used by identifying the allele of a DNA marker and checking if plants have certain genes or QTL instead of phenotypic screening. There are several advantages of using MAS over using phenotypic selection in conventional breeding, including its simplicity and flexibility compared to phenotypic screening, early selection that can be done in seedling stage, high reliability, and independence of selection environments. Effectiveness and efficiency of MAS can accelerate the breeding process. It is useful a technique that can eliminate the issues of time consuming or labor intensive processes such as phenotypic evaluation and instead advances the breeding process quickly. It also gives the advantage of combining multiple genes simultaneously known as "gene pyramiding". Selecting for low heritability traits and traits where phenotypic evaluation is not available makes MAS a very compatible option in crop breeding.

Project objectives

The soyfood market has been growing and expanding in many countries. Seeds with undesirable texture cause difficulties during processing. Stone seeds are known for their complications during soyfood processing because they do not absorb the water. The role of seed hardness is crucial in quality attributes for food-grade soybeans because it affects water absorption, seed coat permeability, and overall texture. Testing for seed hardness is expensive and time consuming and not available for breeders' selection at early stage. Although seed hardness is one of the main traits of food grade soybeans, there is not enough information available on the genetics of the trait.

Calcium content in the seed coat is strongly correlated with water absorption. High calcium content causes harder seeds. The correlation between calcium content and seed hardness was identified yet it was affected by environmental factors (Chen et al., 2001). Hard seed issues

can be solved by using varieties with low calcium content however, not enough genetic information is available on calcium content of soybean.

In soybean, DNA markers are mainly used in identification of QTL for major traits such as protein and oil. For food-grade soybeans, it is important to understand the genetic control of seed hardness and calcium content. Knowing this information would help to manipulate these important seed quality traits and facilitate the process of breeding specialty soybeans for the soyfood market. Therefore, the objectives of this study were to identify SSR markers and QTL for seed hardness and calcium content of soybean and to confirm identified SSR markers and QTL in different genetic backgrounds. Utilization of the markers will improve breeding efficiency via MAS and identify correlation between these two traits.

REFERNCES

Anderson, R. L. and W.J. Wolf. 1995 Compositional changes in trypsin inhibitors, phytic acid, saponins and isoflavones related to soybean processing. J. Nutr. 125: 581S–588S.

Argel, P.J., and C.J. Patron. 1999. Overcoming legume hardseededness. Forage seed production. 2:247-259.

Barnes, S. 2004. Fifth Internat'l Symposium on the role of soy in preventing and treating chronic disease. Soy isoflavones—phytoestrogens and what else? J. Nutr. 134:1225S-1228S.

Basten, C.J., B.S. Weir, and Z.B. Zeng. 1999. QTL cartographyer, version 1.13. Dep. of Statistics, North Carolina State Univ., Raleigh.

Beleia, A., L.T. Thao, and E.I. Ida. 1993. Lowering phytic phosphorus by hydration of soybeans. J. Food Sci. 58:375-388.

Cakmak, I. 2002. Plant Nutrition research: Priorities to meet human needs for food in sustainable ways. Plant and Soil 247: 3-24.

Calero, E., S.H. West, and K. Hinson. 1981. Water absorption of soybean seeds and associated casual factors. Crop Sci. 21:926-933.

Campbell, C.R. and C.O. Plank. 1991. Sample preparation in plant analysis reference procedures for the southern region of the United States. Southern Coop. Serv. Bull. 368:1–11.

Carmona, A., L. Borgudd, G. Borges, and A. Levy-Benshimol. 1996. Effect of black bean tannins on in vitro carbohydrate digestion and absorption. J. Nutri. Biochem. 7(8):445-450.

Cassidy, A., S. Bingham, and K. Setchell. 1995. Biological effects of isoflavonoids in young women: Importance of the chemical composition of soyabean products. Br. J. Nutr. 74:587-6.

Centeno, V., G.D. Barboza, A. Marchionatti, V. Rodriguez, and N.T. Talamoni. 2009. Molecular mechanisms triggered by low calcium diets. Nutr. Res. Rev. 22:163-174.

Chamchalow, N., E.A. Kueneman, and P.A. Nicks. 1993. FAO Institutional Report. Pp. 164-172. In N. Chamchalow and N.L. Laosuwan (eds.). Soybeans in Asia. NH Science publications, Lebanon, NH.

Chen, P., G.R. Buss, and K.C. Diehl. 1993. Physical and chemical characteristics associated with hardness of small-seeded soybean for natto. ASA-CSSA-SSSA International Annual Meetings. Madison, WI.

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

Clarkson, D.T. 1984. Calcium transport between tissues and its distribution in the plant. Plant, Cell, and Environment. 7:449-456.

Cornelious, B., P. Chen, Y. Chen, D. Leon, J.G. Shannon, and D. Wang. 2005. Identification of QTLs underlying water-logging tolerance in soybean. Mol. Breed. 16:103–112.

Cregan, P.B., T. Jarvik, A.L. Bush, R.C. Shoemaker, K.G. Lark, A.L. Kahler, N. Kaya, T.T. VanToai, D.G. Lohnes, J. Chung, and J.E. Specht. 1999. An integrated genetic linkage map of the soybean genome. Crop Sci. 39: 1464-1490.

Dayod, M., S.D. Tyerman, R.A. Leigh, and M. Gilliham. 2010. Calcium storage in plants and the implications for calcium biofortification. Protoplasma 247:215-231.

Dhaubhadel, S., K. Kuflu, M.C. Romero, and M. Gijzen. 2005. A soybean seed protein with carboxylate-binding activity. J. Exp. Bot. 56:2335-2344.

Edwards, J.D., V.M. Lee, and S.R. McCouch. 2004. Sources and predictors of resolvable indel polymorphism assessed using rice as a model. Mol. Genet. Genom. 271: 298–307.

Edwards, K.J., J.H.A. Barker, A. Daly, C. Jones, and A. Karp. 1996. Microsatellite libraries enriched for several microsatellite sequences in plants. Biotechniques 20:758-760.

Fageria, N.K., V.C. Baligar, and C.A. Jones. 1997. Growth and mineral nutrition of field crops. Books in soils, plants, and the environment. p.57. New York: Marcel Dekker.

Fehr, W.R., C.E. Caviness, D.T. Burmood, and J.S. Pennington. 1971. Stage of development descriptions for soybeans, *Glycine Max* (L.) Merill. Crop Sci. 11:929-931.

Geater, C.W. and W.R. Fehr. 2000. Association of total sugar content with other seed traits of diverse soybean cultivars. Crop Sci. 40:1552-1555.

Geater, C.W., W.R. Fehr, and L.A. Wilson. 2000. Association of soybean seed traits with physiological properties of natto. Crop Sci. 40:1529-1534.

Gijzen, M., C. Weng, K. Kuflu, L. Woodrow, K. Yu, and K. Poysa. 2003. Soybean seed luster phenotype and surface protein co-segregate and map to linkage group E. Genome, 46:659-664.

Gijzen, M., K. Kuflu, and P. Moy. 2006. Gene amplification of the *Hps* locus in *Glycine max*. BMP Plant Biol. 6:6

Guillon, F., and M. Champ. 2002. Carbohydrate fractions of legumes: Usses in human nutrition and potential for health. Br. J. Nutr. 88:293:306.

Harada, K. 2001**.** An informative linkage map of soybean reveals QTLs for flowering time, leaflet morphology and regions of segregation distortion. DNA Res 8(2):61-72.

Harland, B.F. and D Oberleas. 1986. Anion exchange method for determination of phytate in foods: Collaborative study. J. Assoc. of Anal. Chem. 69:667-670.

Heaney, R.P., C.M. Weaver, and M.L. Fitzsimmons. 1991. Soybean phytate content: Effect on calcium absorption. J. Clin. Nutr. 53:745-747.

Hosoi, T. and K. Kiuchi. 2003. Natto-a food made by fermenting cooked soybeans with *Bacillus subtilis* (*natto*). p. 264-322. In: E.R. Farnworth (ed.) Handbook of fermented functional foods. CRC Press, Boca Raton, FL.

Hou, F.F. and F.S. Thseng. 1991. Studies on the flooding tolerance of soybean seed: Varietal differences. Euph. 57:169–173.

Hyten, D.L., V.R. Pantalone, C.E. Sams, A.M. Saxton, D. Landau-Ellis, T.R. Stefaniak, and M.E. Schmidt. 2004. Seed quality QTL in a prominent soybean population. Theor. Appl. Genet. 109: 552–561.

Ilarsan, H., R.G. Palmer, J. Imsande, and H.T Horner. 1997. Quantitative determination of calcium oxalate and oxalate in developing seeds of soybean (*Leguminosae*). Am. J. Bot. 84:1042-1046.

Ilarslan, H., R.G. Palmer, and H.T. Horner. 2001. Calcium oxalate crystals in developing seeds of soybean. Ann. Bot. 88:243-257.

Johnson, D. 2000. Edamame. Engineering and technology for a sustainable world. 7:11-12.

Johnson, H.W. and R.L. Bernard. 1962. Soybean genetics and breeding. Adv. Agron. 14: 149-221.

Kalkwarf, H.J., J.C. Khoury, and B.P. Lanphear. 2003. Milk intake during childhood and adolescence, adult bone density, and osteoporotic fractures in US women. Am. J. Clin. Nutr. 77:10-11.

Kannenberg, L.W. and R.W. Allard. 1960. An association between pigment and lignin formation in the seed coat of the Lima bean. Crop Sci. 4:621-622.

Katsuyama, H., S. Ideguchi, M. Fukunaga, K. Saijoh, and S. Sunami. 2002. Usual dietary intake of fermented soybeans (*natto*) is associated with bone mineral density in premenopausal women. J. Nutr. Sci. Vitaminol (Tokyo). 48:207–15.

Kearsey, M.J. 2002. QTL analysis: Problems and (possible) Solutions. pp. 45-58. In M.S. Kang (ed). Quantitative genetics, genomics, and plant breeding. CABI Pub. New York.

Keim, P., B.W. Diers, and R.C. Shoemaker. 1990. Genetic analysis of soybean hard seededness with molecular markers. Theor. Appl. Genet. 79:465-469.

Keiser, J.R. and R.E. Mullen. 1993. Calcium and relative humidity effects on soybean seed nutrition and seed quality. Crop Sci. 33:1345-1349.

Kilen, T. and E. Hartwig. 1978. An inheritance study of impermeable seed in soybeans. Field Crops Res. 1:65–70.

Kinzel, H. 1989. Calcium in the vacuoles and cell walls of plant tissue. Flora. 182:99–125.

Kisha, T., C.H. Sneller, and B.W. Diers. 1997. Relationship between genetic distance among parents and genetic variance in populations of soybean. Crop Sci. 37:1317–1325.

Kohyama, K., M. Murata, F. Tani, Y. Sano, and E. Doi. 1995. Effects of protein composition of gelation of mixtures containing soybean 7s and 11s globulins. Biosci. Biotech. Biochem. 59(2):240-245.

Korzun, V. 2002. Use of Molecular Markers in Cereal Breeding. Cell. Mol. Bio. Letters. 7:811 – 820.

Kosambi, D.D. 1944. The estimation of map distance from recombination values. Ann. Eugen. 12:172–175.

Liu, B., T. Fujita, Z. Yang, S. Sakamoto, D. Xu, and J. Abe. 2007. QTL Mapping of domestication-related traits in soybean (*Glycine max*). Ann. Bot. 100: 1027–1038.

Liu, K. 1997. Agronomic characteristics, production, and marketing. pp. 1-24. In K. Lui (ed.), Soybeans: Chemistry, technology, and utilization. Chapmen & Hall, New York, NY.

Lucas, E.W., D.R. Erickson, and W.K. Nip. 1989. Food uses of whole oil and protein seeds. pp 40-87. In S. Chen (ed.), Principles of soymilk production. Illinois: American Oil Chemist's Society.

Lui, K.S. 1996. Immature soybeans. Direct use for food. Inform. 7:1217-1223.

Lumpkin, T.A., J.C. Konovsky, K.J. Larson, and D.C. McClary. 1993. Potential new specialty crops from Asia: Azuki bean, edamame soybean, and astragalus. p. 45–51. In: J. Janick and J.E. Simon (eds.), New crops. Wiley, New York.

Ma, F., E. Cholewa, T. Mohamed, C.A. Peterson, and M. Gijzen. 2004. Cracks in the palisade cuticle of soybean seed coat correlate with their permeability to water. Ann. Bot. 94:213-228.

Ma, G., Y. Jin, J. Piao, F. Kok, B. Guusje, and E. Jacobsen. 2005. Phytate, calcium, and zinc contents and their molar ratios in foods commonly consumed in China. J. Agric. Chem. 53:10285-10290.

Mackay, T.F.C. 2001. The Genetic architecture of quantitative traits. Ann. Rev. Genet. 35:303-39.

Mason, A.C., C.M. Weaver, S. Kimmel, and R.K. Brown. 1993. Effect of soybean phytate content on calcium bioavailability in mature and immature rats. J. Agric Food Chem. 41:246-249.

Masuda, R. 1991. Vegetable soybean research needs for production and quality improvement. Asian Veget. Res. Dev. Cent., pp92-102.

Masuda, R., K. Hashizume, and K. Kaneko, 1988. Effect of holding time before freezing on constituents and flavour of frozen green soyabeans. J. Jap. Soc. Food. Sci. Tech. 35(11): 763-770.

Messina, M.J. and C.L. Loprinzi. 2001**.** Soy for Breast Cancer Survivors: A Critical Review of the Literature. J. Nutr. 131: 3095S-3108.

Miles, C.A., T.A. Lumpkin, and L. Zenz. 2000. Edamame. Washington State University Cooperative Extension Bulletin No. PNW0525.

Min, S., Y.Yu, and S. Martin. 2005. Effect of soybean varieties and growing locations on the physical and chemical properties of soymilk and tofu. J. Food Sci. 70:C8-C12.

Minarik, C.E., and J.W. Shive. 1939. The effect of boron in the substrate on calcium accumulation by soybean plants. J. Bot. 26:827-831.

Minorsky, P. V. 1985. A heuristic hypothesis of chilling injury in plants: A role for calcium as the primary physiological transducer of injury. Plant, Cell, and Environment 8:75-94.

Mullin, W.J. and W. Xu. 2001. Study of soybean seed coat components and their relationship to water absorption. J. Agri. Food. Chem. 49 (11): 5331-5335.

Norman A.G. 1963. The Soybean Genetics Breeding Physiology Nutrition Management. Academic Press. New York and London.

Potts, H.C., J. Duangpatra, W.G. Hairston, and J.C. Delouche. 1978. Some influences of hardseededness on soybean seed quality. Crop Sci 18: 221-224.

Poysa, V., L. Woodrow, and Y. Yu. 2006. Effect of soy protein subunit composition on tofu quality. Food res. Int. 39:309-317.

Qutob, D., F. Ma, C.A. Peterson, M.A. Bernards, and M. Gijzen. 2008. Structural and permeability properties of the soybean seed coat. Botany 86: (3) 219-227.

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

Rengel, Z. 1992. Role of calcium in aluminum toxicity. New Phytol. 121:499-513.

Saio, K., K. Arai, and T. Watanabe. 1973. Fine structure of soybean seed coat and its changes on cooking. Cereal Sci. Today 18:197–201.

Sakamoto, S., J. Abe, A. Kanazawa, and Y. Shimamoto. 2004. Marker-assisted analysis for soybean hard seededness with isozyme and simple sequence repeat loci. Breeding Science. 54(2):133-140.

Sankaran, R.P., T. Huguet, and M.A. Grusak. 2009. Identification of QTL affecting seed mineral concentration and content in the model legume *Medicago Truncatula*. Theor. Appl. Genet. 119(2): 241-253.

Shanmugasundaram, S., C.S. Tsou, and S.H. Cheng. 1989. pp.1987-1987. In A.J. Pascade (ed.), Vegetable soybean in the east. World Soybean Research Conference IV.

Shao, S., C.J. Meyer, F. Ma, C.A. Peterson, and M.A. Bernards. 2007. The outermost cuticle of soybean seeds: Chemical composition and function during imbibitions. J. Exp. Bot. 58:1071-1082.

Sharp, P.J., S. Johnston, G. Brown, R.A. Mclntosh, M. Pallotta, M. Carter, H.S. Bariana, S. Khatkar, E.S. Lagudah, R.P. Singh, M. Khairallah, R. Potter, and M.K. Jones. 2001. Validation of molecular markers for wheat breeding. J. of Agr. Res. 52(12):1357-1366.

Shu, X.O., F. Jin, Q. Dai, W. Wen, J.D. Potter, L.H. Kushi, Z. Ruan, Y.T. Gao, and W. Zheng. 2001. Soyfood intake during adolescence and subsequent risk of breast cancer among Chinese women. Cancer Epidemiol Biomarkers Prev. 10:483–488.

Smith, K.J., and W. Huyser. 1987. World distribution and significance of soybean. In: J. R. Wilcox (ed.), Soybeans: improvement, production, and uses, 2nd edn. Agronomy 16:23-48.

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

Taira, H. 1990. Quality of soybeans for processed foods in Japan. Japan Agric. Res. Quarterly 24:224-230.

Taira, H., H. Takahashi, H. Okano, and S. Nagashima. 1983. Suitability of Soybean cultivars grown in Ibaraki prefecture for natto production. In quality of soybean seeds grown in Japan. Report national food research institute 50:62-71.

Todd, J.J. and L.O. Vodkin. 1993. Pigmented soybean (*Glycine max*) seed coats accumulate proanthocyanidins during development. Plant Physiol. 102:663-670.

Tyler, J.M. 1997. Effect of impermeable seed coat on germination of seed for early maturing soybean. Seed Technology 19:45-50.

Van Ooijen, J.W., and R.E. Voorrips. 2001. JoinMap version 3.0, software for the calculation of genetic linkage maps. Plants Research Int., Wageningen, the Netherlands.

Voorrips, R.E. 2002. Mapchart: Software for the graphical presentation of linkage maps and QTLs. J. Hered 93:77–78.

Wang, D., J. Shi, S.R. Carlson, P.B. Cregan, R.W. Ward, and B.W. Diers. 2003. A low-cost, high-throughput polyacrylamide gel electrophoresis system for genotyping with microsatellite DNA markers. Crop Sci. 43:1828–1832.

Wang, H.L., E.W. Swain, and W.F. Kwolek. 1983. Effect of soybean varieties on the yield and quality of tofu. Cereal Chem. 60(3):245-248.

Weaver, C.M., R.W. Proulx, and R. Heaney**.** 1999. Choices for achieving adequate dietary calcium with a vegetarian diet. Am. J. Clin. Nutr. 70:543S-548S.

Werker, E., I. Marbach, and A. Mayer. 1979. Relation between the anatomy of the testa, water permeability and the presence of phenolics in the genus *Pisum*. Ann Bot 43**:** 765-771.

White, P.J. and M.R. Broadley. 2003. Calcium in plants. Ann. Bot. 92:487-511.

World Health Organization. 2008. Available at http://www.who.int/en/; verified 2 May 2009. WHO, Geneva, Switzerland.

Yoshikawa, Y., T. Kawashima, T. Ishibashi and P. Chen. 2008. Associations of soybean seed quality attributes with sensory properties of natto. Joint annual meeting. 636-19. 5-9 Oct, 2008. Houston, TX.

Yalkich, R.W., W.P. Wergin, and E.L. Vigil. 1986. Special secretory cells in the soybean seed coat. Protoplasma 134:78-87.

Yamanaka, N., S. Ninomiya, M. Hoshi, Y. Tsubokura, M. Yano, Y. Nagamura, T. Sasaki, K. Young, V.R., and P.L. Pellett. 1994. Plant proteins in relation to human protein and amino acid nutrition. Am. J. Clin. Nutr. 59(s):1203S–12S.

Yukihiro, I., I. Masayuki, M. Akemi, K. Etsuko, K. Sadanobu, K. Yoshiko, and H.Yoneshima. 2006. Intake of fermented soybeans, *Natto*, is associated with reduced bone loss in postmenopausal women: Japanese Population-Based Osteoporosis (JPOS). J. Nutr. 136:1323-1328.

Zabala, G. and L. Vodkin. 2003. Cloning of the pleiotropic *T* locus in soybean and two recessive alleles that differentially affect structure and expression of the encoded flavonoid 3'hydroxylase. Genetics 163:295-309.

Zhang, W.K., Y.J. Wang, G.Z. Luo, J.S. Zhang, C.Y. He, X.L.Wu, J.Y. Gai, and S.Y. Chen. 2004. QTL mapping of ten agronomic traits on the soybean (G*lycine max* L. Merr.) genetic map and their association with EST markers. Theor. Appl. Genet. 108:1131–1139.

Zhang. B., P. Chen, C.Y. Chen, D. Wang, A. Shi, A. Hou, and T. Ishibashi. 2008. Quantitative trait loci mapping of seed hardness in soybean. Crop Sci. 48:1341–134.

Zhang, B. (2006). Genetic analysis of food-grade soybeans and identification of quantitative trait loci (QTL) for seed hardness and calcium content in soybean.

Chapter II

Putative Quantitative Trait Loci (QTL) Associated with Seed Calcium Content in Soybean

ABSTRACT

Soybean [*Glycine max* (L.) Merr.] seed calcium content is an important trait for soyfood production as one of the main determining characteristics of seed texture affecting soyfood quality. Seeds with higher calcium content have harder seeds which is not desirable for food-grade soybean varieties such as natto. Seed calcium content analysis is time consuming and labor intensive. Molecular markers and QTL for seed calcium content will facilitate the development of elite cultivars with desired calcium content through marker assisted selection. The objectives of this study were to identify new QTL/markers associated with seed calcium content and confirm previously reported QTL for seed calcium content. The crosses from KS4303sp (low calcium) x PI 407818 B (high calcium) and PI 408052 C (low calcium) x PI 407818 B (low calcium) with 120 and 124 RIL populations, respectively, were screened with SSR markers and tested for calcium content. The $F_{2:4}$ and $F_{2:5}$ lines from the two populations were grown in a randomized complete block design (RSBD) with two replications, at Fayetteville, AR and Keiser, AR in 2008 and 2009. Composite interval mapping for the SSR markers revealed a new putative QTL *Ca5* for seed calcium content that resides on an approximate 13cM region and linked to markers Sat_290 and Satt115 on chromosome 18. Single marker analysis revealed markers stable across environments such as Satt184 and Satt077 on chromosomes 1, Satt095 on chromosome 2, Sat_356 on chromosome 5, Satt175 on chromosome 7, Sat_377 on chromosome 8, Satt217, Sat_290, Satt115, and Sat_358 on chromosome 18, and Sat_174 on chromosome 20 for seed calcium content. Previously reported QTL *Ca1* and *Ca2* on chromosomes 8 and 20, respectively, were confirmed in present study. Confirmed QTL/markers reported previously and new QTL/markers identified in the present study offer a potential tool for MAS for calcium content. Findings of this study will facilitate development of food-grade soybean varieties with low calcium content for natto production.

INTRODUCTION

Soybean was first domesticated in China and later spread to Korea, Japan and other countries in Asia. Only after 1700th, soybean was introduced to the US. Initially, soybean was grown in the US as a forage crop and it has been grown as a grain crop since 20th century. Although soyfood has been popular in Asia since ancient times it was not introduced to the US until later. Soybeans are now becoming more popular in Western diet due to their nutritional values and health benefits. In order to meet consumer requirements, it is very important to have soybean varieties with specialty quality traits that are suitable for certain types of soyfood. Depending on specifics of processing for different end uses, soyfood may require small or large seeds with certain colors or with some other traits. Soybean sprouts and natto require small seeded soybeans and they should have high water absorbability. During the process of soaking seeds should absorb water, but often seeds with low water absorption cause difficulty. It is important to have uniform seeds for food grade soybeans. Low water absorbability causes stone seeds, which is not desirable because they lower germination rate and seedling vigor qualities of soybean sprout and affect the texture of natto (Mullin and Xu, 2001). Removing the stone seeds will cost extra (Mullins and Xu, 2001) and affects the quality of the end product. For soyfood product such as tofu and natto, the texture is very important. Texture is affected by many quality traits such as water absorption, seed coat permeability, calcium content, and cooked seed hardness.

Calcium content in the seed coat affects its water absorption. Chen et al., (2001) indicated that calcium content in the seed was positively correlated with seed hardness after cooking and both traits varied due to environmental effects such as temperature and soil type. Since calcium has a significant role in the cell membrane, it could affect water absorption by making the cell wall thicker or having calcium oxalate cell layer on soybean seed wall.

Calcium has an important effect on soyfood quality and requires thorough genetic investigation. Information about its genetic control and interaction with the environment will accelerate the process of developing cultivars with desired calcium content in the seed. With the advance of molecular techniques, it is easier to manipulate quantitative traits using molecular markers associated with QTL. QTL for calcium content has been mostly researched on common beans (*Phaseolus vulagris* L.). Guzman et al., (2003) found two QTL associated with calcium content on common beans. They also found that these QTL had close association with other mineral contents. Klein and Grusak (2009) identified two QTL for calcium content in the model legume *Lotus japonicus* on chromosomes 1 and 2. They found a negative correlation between calcium content and seed mass, which is similar to what Guzman et al., (2003) found in common beans. In the same study Klein and Gursak (2009) suggested that calcium movement into seeds may be influenced by genetic control. In soybean, Zhang et al., (2009) found four QTL, *Ca1-4*, from chromosomes 7, 8, and 20 for calcium content. They also found that QTL *Ca1* on chromosome 8 for calcium content shared the same genetic region with a seed size QTL. However, these QTL have not been confirmed in different genetic backgrounds.

Having genetic control over a trait such as calcium content would ease the process of calcium content analysis which is time consuming and labor intensive. As the genetics of this trait applies to specialty soybeans for soyfood production, molecular markers and QTL, once identified and confirmed, can be used to improve the micronutrient concentration to obtain high calcium varieties for nutrient considerations or varieties with low calcium where texture of the product is important. QTL for calcium content of soybean will be a useful tool to facilitate the development of elite cultivars with desired calcium amount through marker assisted selection (MAS).

The objectives of this study were to identify SSR markers and novel QTL for calcium content of soybean seed in plant populations derived from low x high calcium crosses and to confirm previously reported QTL.

MATERIALS AND METHODS

Population development and field experiment

A high calcium soybean line PI 407818 B was crossed with low calcium lines KS4303sp and PI 408252 C. The crosses were made in spring of 2007 in the greenhouse. The F_1 plants were space planted in the field at the Arkansas Research and Extension Center affiliated with the University of Arkansas in Fayetteville, AR in summer 2007. Flower color and leaf shape were used as morphological marker to identify true hybrids. F_2 plants were space planted and harvested individually in a winter nursery in Costa Rica in winter of 2007 and 100 seeds from each F_2 plant was taken to obtain next generation, $F_{2:3}$ lines. Each $F_{2:3}$ line was grown in 3 m row and each row was bulk harvested in the Costa Rica winter nursery in spring 2008. A total of 200 $F_{2:3}$ lines were developed from each cross KS4303sp X PI 407818 B and PI 408052 C X PI 407818 B. One hundred twenty progeny lines from the cross KS4303sp X PI 407818 B and 124 progeny lines from the cross PI 408052 C X PI 407818 B with adequate amount of seeds were selected as a source of populations for QTL mapping. In summer 2008, $F_{2:4}$ lines from both populations were grown in the field with complete randomized block design (RCBD) with two replications at two locations Fayetteville, AR and Keiser, AR. Parental lines KS4303sp, PI 407818 B, and PI 408052 C were included as checks in the test. Each line was grown in a single row plot with 3 m in length and 0.95 m row spacing. Irrigation was conducted according to University of Arkansas irrigation schedule. Once plants in a row reached 95% maturity, they were bulk harvested either with a combine or a plant thresher. Field plots were managed using

standard cultural practices adopted for full season soybean production in Arkansas. In 2009, $F_{2:5}$ lines were grown in RCBD with two replications at two locations; Fayetteville, AR and Keiser, AR. Plots were managed in the same way as described for $F_{2:4}$ lines.

DNA extraction

Leaf tissues were collected from the uppermost partly developed trifoliate leaf of every plant of each $F_{2:4}$ line. The bulked tissue samples were ground with a mortar and pestle in liquid nitrogen and stored in a -20^{o}C freezer. Total genomic DNA was isolated using the CTAB (hexadercyltrimethyl ammonium bromide) method (Kisha et al., 1997). Briefly, 0.50 mg powder obtained from ground leaves were transferred to a 2.0 ml microtube and 750 µl of CTAB extraction buffer was added followed by incubation at 65^{o}C for 60 min. with inverting every 10 min. Tubes were cooled and 500 µl chloroform:isoamyl alcohol (24:1) was added and inverted several times followed by centrifuging samples at 13,500 for 15 min. The aqueous layer was transferred to a 2.0 ml microtube and centrifuged for 5 min. which lead to DNA precipitation. DNA was washed with 95% ethanol solution. Supernatant was poured out and DNA pellets were allowed to air dry at room temperature overnight. 250 µl of 0.1 x TE was added over dried DNA and left in temperature for a day to dissolve DNA. Each sample was checked for its concentration using Bio-Tek PowerWave XS Microplate Spectrophotometer.

SSR polymorphism analysis

Polymerase chain reaction (PCR) and simple sequence repeat (SSR) primer genotyping were performed in a 96-well or 384-well BIO-RAD iCycler thermocycler (Bio-Rad, Hercules, CA). PCR was performed using 4 µl of 20ng/µl template DNA, 1.0 µM of 0.5 µM forward and reverse primers, 0.2 µl of Go Taq Flexi polymerase, 0.9 µl of 2.5 mM dNTP (nucleotides) mixture, 1.8 µl of 2.5 mM $MgCl_2$, and 4.3 deionized H_2O. PCR reaction were at 94^{o}C for 4 min.

followed by 33 cycles at 94^{o}C for 25 s for denaturation, 47^{o}C for 25 s for annealing, 68^{o}C for 25 s for extension, and 72^{o}C for 5 min. for final extension after the last cycle. Amplified PCR products were separated on a 6% non-denaturing polyacrylamide gel on a Megagel electrophoresis system (C.B.S. Scientific) and stained with ethidium bromide (Wang et al., 2003). The gels were visualized under UV fluorescence and digital images were taken to score individuals for marker alleles. Parental genotypes, KS4303sp, PI 407818 B, and PI 408052 C were screened using 486 SSR markers from 20 chromosomes of soybean. Markers scored as monomorphic when both parents exhibited amplification products of the same size and as polymorphic when parents had fragments in different size. Only polymorphic markers were chosen to screen the progenies. There were 105 polymorphic markers for the cross KS4303sp X PI 407818 B and 101 polymorphic markers for the cross PI 408052 C X PI 407818 B. Primers for the SSR markers used in this study, their sequences, and their integration in soybean molecular linkage map are publicly available online at http://129.186.26.94/ssr.html (Cregan et al., 1999, Song et al., 2004).

Calcium analysis

Calcium content was determined by the HNO_3 method (Campbell and Plank, 1991). Ten grams of seeds from each plot were ground using a Krups KM75 coffee grinder until fine powder was obtained. 0.25 grams of powder were digested by 2.5 ml HNO_3. Samples were gradually heated to 60^{o}C for 45 min. and slowly increased to 120^{o}C for an hour after adding 1 ml H_2O_2. Cooled samples were mixed with deionized water and filtered through #41 quantitative paper followed by calcium content analysis using Spectro Ciros with simultaneous inductively coupled plasma (Spectro Analaytical Instruments, Inc., Mahwah, NJ).

Linkage map construction and QTL analysis

Association between molecular data and phenotypic data used to initially identify candidate QTL was analyzed by single-factor analysis of variance with the PROC GLM procedures of SAS software v9.2 (SAS Institute, 2002). Linkage mapping and QTL analysis procedures were performed by using JoinMap 3.0 (Van Ooijen and Voopris, 2001) with Kosambi function (Kosambi, 1944) and WinQTL cartographer (Basten et al., 1999) with single marker analysis (SMA) and interval mapping (IM) to identify QTL associated with calcium content. For SMA, $p<0.05$ was used as a threshold for significant markers (Cornelious et al., 2005). In IM, the empirical significance threshold was determined by 1000 permutations, with a walk speed of 1 cM, and a significance level of alpha =0.05 (Cornelious et al., 2005). Mapchart (Voopris, 2002) was used to create the LOD plots according to the data from WinQTL Cartographer.

Statistical data analysis

Phenotypic data for seed calcium content collected in 2008 and 2009 were both analyzed separately and combined using the PROC MEANS procedure in SAS software v9.2 (SAS Institute, 2002). The PROC GLM procedure was used to determine variance components for calcium content with location and replication as random blocking factors. Distribution of calcium content in two locations and in two years were analyzed using JMP software.

RESULTS AND DISCUSSION

Table 1 and 2 present the analysis of variance for calcium content in 120 RIL from the cross KS4303sp x PI 407818 B (population 1) and 124 RIL from the cross PI 408052 C x PI 407818 B (population 2) grown at two locations over two years. The model was significant with p-value of <0.0001 for both populations and R^2 values were 0.92 and 0.87 in population 1 and 2,

respectively. The model was based on the differences for the following sources of variation: year, location, year x location, genotype, genotype x year, genotype x location, and year x location x genotype. Differences in all sources were statistically significant in population 1 derived from KS43030sp X PI 407818 B. Differences in all sources except for location x genotype and year x location x genotype were statistically significant inpopulation 2 derived from PI 408052 C x PI 407818 B which allowed combined data across locations and years to be used in analysis.

ANOVA from both populations also showed that there were large effects of location, year, and location x year interaction on calcium content. Therefore, genotypes need to be evaluated across multiple environments for this trait in a practical breeding program.

Figures 1-4 show the frequency distribution of calcium content at two locations in two years from population 1. Figures 5-8 show the frequency distribution of calcium content at two locations and in two years from population 2. These data show normal distribution of calcium content over years and across locations in both populations (Figures 9-12). Calcium content in population 1 ranged between 0.18- 0.29% at Fayetteville and 0.2-0.31% at Keiser in 2008, 0.18-0.30% at Fayetteville and 0.21-0.33% at Keiser in 2009 and low calcium parent KS4303sp had 0.18% and high calcium parent PI 407818 B had 0.32% calcium in average. Calcium content in population 2 ranged between 0.2-0.32% at all locations and years except for at Keiser in 2009 where the range of calcium content was between 0.22-0.35%. In population 2, low calcium parent PI 408052 C had 0.22% calcium content while high calcium parent PI 407818 B had 0.32% on average. Higher calcium content was observed at Keiser in both years in both populations indicating the attribution to the differences in environmental factors associated with each location. A significant difference in calcium content in 2008 and 2009 was likely explained

by the weather conditions in each year. 2009 was rainy and colder than 2008 and as a consequence harvesting time shifted to a much later date. Also higher calcium content in RIL may be a result of plant reaction to stress that was caused with weather conditions in 2009. This difference was mostly in magnitude rather than change in rank of the RIL in the whole population. Since magnitude shift of calcium content was found between years and locations, QTL analysis was performed for each year and location separately prior to combined data analysis.

Out of total 486 SSR markers screened, 105 SSR markers were polymorphic between the parental lines, KS4303sp and PI 407818 B (Table 3). These polymorphic markers were distributed in 19 of 20 chromosomes of soybean and they were used to screen the 120 RIL population. A larger numbers of polymorphic markers were found in chromosomes 1, 2, 7, 8, 13, and 18. On average, the density of polymorphic markers ranged from two to nine per chromosome with an average of one marker per 12.5 cM. Out of 486 SSR markers, 101 markers were found to be polymorphic between parental lines PI 407818 B and PI 408052 C (Table 4). These polymorphic markers were distributed one every 12.4 cM in average in all 20 chromosomes of soybean and they were chosen to screen the 124 RIL population. Chromosomes 1, 2, and 18 had a larger number of polymorphic markers. Prior to linkage analysis, each polymorphic marker was analyzed for the goodness-of-fit the expected allelic segregation ratio (1:2:1) in the population.

Single marker analysis in each year and each location in population 1 revealed three markers on chromosome 1, two markers on chromosome 2, three markers each on chromosomes 7 and 8, one marker each on chromosomes 10, 13, and 16, three markers on chromosome 18, and one marker on chromosome 20 in 2008 (Table 5). In 2009, there were more markers associated

with the trait; five markers on chromosome 1, four markers on chromosome 2, three markers each on chromosomes 7 and 8, one marker on chromosome 9, two markers each on chromosomes 10 and 13, one marker each on chromosomes 16 and 17, and three markers each on chromosomes 18 and 20 (Table 5). The lower number of markers in 2008 may be due to some missing data due to poor germination of seeds from winter nursery. A total of 11 chromosomal regions contained markers associated with seed calcium content. Some of the markers were associated with seed calcium content in all environments, hence found to be most important. For example, Satt184 on chromosome 1, Satt095 on chromosome 2, Satt175 on chromosome 7, Sat_377 on chromosome 8, Satt123 on chromosome 10, Satt547 on chromosome 16, and Satt174 on chromosome 20. These markers contributed from 10 to 34% of calcium variation in a 120 RIL population derived from KS43030sp x PI 407818 B. Among these stable calcium markers, Sat_377 on chromosome 7 and Sat_174 on chromosome 20 were reported to be linked to calcium QTL *Ca1* and *Ca2* (Zhang et al., 2009), respectively. Confirming QTL is an important step for MAS, therefore, these regions will be important for seed calcium selection.

Eleven chromosomal regions also contained markers that were stable in three of four environments such as Sat_353 and Satt254 on chromosome 1, Satt677 and Sat_288 on chromosome 7, Sat_199 and Satt377 on chromosome 8, and Satt656 on chromosome 13. These markers contributed to calcium content in range from 7 to 14% in the KS4303sp x PI 407818 B population. Other markers on chromosomes 2, 9, 18, and 20 contained markers associated with seed calcium content, yet were not stable across environments and may need more investigation in future.

Single marker analysis in each year and each location in population 2 derived from PI 408052 C x PI 407818 B revealed two markers each on chromosomes 1 and 5, three markers on

chromosome 8, two markers on chromosome 10, one marker each on chromosomes 12 and 17, and four markers on chromosome 18 to be associated with calcium content in 2008. In 2009, three markers on chromosome 1, two markers on chromosome 5, one marker on chromosome 7, three markers on chromosome 3, one marker each on chromosomes 10, 12, 17, and four markers on chromosome 18 to be associated with calcium content (Table 6). Combined data analysis indicated a total of 16 SSR markers from 8 different chromosomes of soybean were associated with calcium content. Most of these markers were stable over years and across locations such as Satt077 on chromosome 1, Sat_356 on chromosome 5, Sat_377 on chromosome 8, Satt541 on chromosome 12, Satt217, Sat_290, Satt115, Sat_358 on chromosome 18. These markers showed significant association with seed calcium content in all locations and the years (Table 6) and contributed to seed calcium content in range of seven to 13%. Sat_377 on chromosome 8 was reported to be linked to calcium QTL *Ca1* (Zhang et al., 2009) and confirmed in this population. The rest of the chromosomal regions were significant in two of three environments, hence require further research as a potential regions containing other QTL for calcium content.

Interval mapping using WinQTL cartographer for the SSR markers revealed a new QTL *Ca5* (proposed) on chromosome 18 linked with markers Sat_290 and Satt115. Alignment of these markers was in agreement with the public map (Song et al., 2004) indicating that the genomic sequence of the PI 408052 C x PI 407818 B is in the same pattern as the original public mapping. This putative calcium QTL *Ca5* resides in an approximately 13 cM region between markers Sat_290 and Satt115 (Figure 13). This new QTL contributed 13% of the phenotypic variation in the PI 408052 C x PI 407818 B population.

Comparison of SSR markers and QTL associated with seed calcium content in two populations derived from KS4303sp x PI 407818 B and PI 408052 C x PI 407818 B revealed

chromosomal regions stable across different environments and genetic backgrounds. Among these chromosomal regions associated with seed calcium content in two populations, markers on chromosome 1, 8, and 18 were more consistent than other chromosomal regions. In addition, Sat_377 on chromosome 8 linked to calcium QTL *Ca1* (Zhang et al., 2009) showed stable association with the trait in both populations and in all environments, hence this calcium QTL has a potential for MAS for calcium.

Comparison of marker allele effects in KS4303sp x PI 407818 B population showed differences ranging from 0.01 to 0.02% calcium content between lines with high calcium parent PI 407818 B allele and lines with low calcium parent KS43030sp allele. Similar results were found in the PI 408052 C x PI 407818 B population. Stable markers in different environments and different genetic backgrounds contributing to more differences such as chromosomal regions on 1, 8, and 18 need more thorough research to investigate the specific contribution of QTL alleles.

In conclusion, confirmed QTL/markers reported previously and new QTL/markers identified in the present study offer a potential tool for MAS for calcium content. Confirmed calcium QTL *Ca1* and *Ca2* will improve efficiency for food-grade soybean varieties for natto production. New QTL *Ca5* and markers associated with calcium content will be confirmed in different environments and genetic backgrounds offering for MAS for calcium content. Findings of this study will facilitate development of food-grade soybean varieties with low calcium content for natto production.

REFERENCES

Basten, C.J., B.S. Weir, and Z.B. Zeng. 1999. QTL cartographer, version 1.13. Dept. of Statistics, North Carolina State Univ., Raleigh.

Campbell, C.R. and C.O. Plank. 1991. Sample preparation in plant analysis reference procedures for the southern region of the United States. Southern Coop. Serv. Bull. 368:1–11.

Chen. P., M.R. Huhn, G.R. Buss, I. Gunduz, and J.H. Wilson. 2001. Interrelationship among seed quality traits of specialty soybeans for natto and tofu. ASA-CSSA-SSSA International annual meeting. Oct. 21-25, Charlotte, NC.

Cornelious, B., P. Chen, Y. Chen, D. Leon, J.G. Shannon, and D. Wang. 2005. Identification of QTLs underlying water-logging tolerance in soybean. Mol. Breed. 16:103–112.

Cregan, P.B., T. Jarvik, A.L. Bush, R.C. Shoemaker, K.G. Lark, A.L. Kahler, N. Kaya, T.T. VanToai, D.G. Lohnes, J. Chung, and J.E. Specht. 1999. An integrated genetic linkage map of the soybean genome. Crop Sci. 39: 1464-1490.

Golbitz, P. 1995. Traditional soyfoods: Processing and products. J. Nutr. 125:570-572.

Guzman-Maldonado, S.H., O. Martiez, J.A. Acosta-Gallegos, F. Guevara-Lara, and O. Peredes-Lopez. 2003. Putative quantitative trait loci for physical and chemical components of common bean. Crop Sci. 43:1029-1035.

Kisha, T., C.H. Sneller, and B.W. Diers. 1997. Relationship between genetic distance among parents and genetic variance in populations of soybean. Crop Sci. 37:1317-1325.

Klein, M.A. and M.A. Grusak. 2009. Identification of nutrient and physical seed trait QTL in the model legume *Lotus japonicus*. Genome 52:677-691.

Kosambi, D.D. 1944. The estimation of map distance from recombination values. Ann. Eugen. 12:172–175.

Li, Y.D., Y.J. Wang, Y.P. Tong, J.G. Gao, J.S. Zhang, and S.Y. Chen. 2005. QTL mapping of phosphorus deficiency tolerance in soybean (*Glycine max* L. Merr.). Euphytica 142(1-2):137-142.

Mullin, W.J. and W. Xu. 2001. Study of soybean seed coat components and their relationship to water absorption. J. Agri. and Food Chem. 49 (11): 5331-5335.

Panthee, D.R., P. Kwanyuen, C.E. Sams, D.R. West, A.M. Saxton, and V.R. Pantalone. 2004. Quantitative trait loci for B-conglycinin (7S) and Glycinin (11S) fractions of soybean storage protein. J. Amer. Oil Chem. Soc. 81(11):1005-1012.

Primomo, V., V. Poysa, G. Ablett, C. Jackson, M. Gijzen, and I. Rajcan. 2005. Mapping QTL for individual and total isoflavone content in soybean seeds. Crop Sci. 45(6):2454-2464.

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

SAS Institue. 2002. SAS/STAT User's guide, Version 9.6 ed. SAS Institute Inc. Cary, NC.

SAS Institute, 2006: JMP User's guide, Version 6.0. SAS Institute Inc., Cary, NC.

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

Stombaugh, S.K., J.H. Orf, H.G. Jung, K. Chase, K.G. Lark, and D.A. Somers. 2004. Quantitative trait loci associated with cell wall polysaccharides in soybean seed. Crop Sci. 44:2101-2106.

Van Ooijen, J.W. and R.E. Voorrips. 2001. JoinMap version 3.0, software for the calculation of genetic linkage maps. Plants Research Int., Wageningen, the Netherlands.

Voorrips, R.E. 2002. Mapchart: Software for the graphical presentation of linkage maps and QTLs. J. Hered. 93:77–78.

Wang, D., J. Shi, S.R. Carlson, P.B. Cregan, R.W. Ward, and B.W. Diers. 2003. A low-cost, high-throughput polyacrylamide gel electrophoresis system for genotyping with microsatellite DNA markers. Crop Sci. 43:1828-1832.

White, P.J. and M.R. Broadley. 2003. Calcium in plants. Ann Bot 92:487-511.

Zhang, B., P. Chen, A. Shi, A. Hou, T. Ishibashi, and D. Wang. 2009. Putative quantitative trait loci associated with calcium content in soybean seed. J. Hered. 100(2):263-269.

Zhang. B., P. Chen, C.Y. Chen, D. Wang, A. Shi, A. Hou, and T. Ishibashi. 2008. Quantitative trait loci mapping of seed hardness in soybean. Crop Sci 48:1341–134.

Table 1. Analysis of variance for soybean seed calcium content of 120 RIL from the cross KS4303sp x PI 407818 B grown at Fayetteville, AR and Keiser, AR in 2008 and 2009.

Source	Degrees of freedom	Sum of squares	Mean square	F-value	P-value	R^2
Model	490	1.09	0.002	9.68	<0.0001	0.92
Year	1	0.09	0.009	385.3	<0.0001	
Location	1	0.37	0.37	1599.74	<0.0001	
Year*Location	1	0.05	0.05	240.32	<0.0001	
Genotype	119	0.42	0.003	15.1	<0.0001	
Year*Genotype	119	0.06	0.0004	2.02	<0.0001	
Location*Genotype	119	0.04	0.0003	1.52	0.001	
Year*Location*Genotype	119	0.04	0.0003	1.4	0.009	
Error	404	0.09	0.0002			

Table 2. Analysis of variance for soybean seed calcium content of 124 RIL from the cross PI 408052 C x PI 407818 B grown in Fayetteville, AR and Keiser, AR in 2008 and 2009.

Source	Degrees of freedom	Sum of squares	Mean square	F-value	P-value	R^2
Model	501	0.98	0.002	6.06	<0.0001	0.87
Year	1	0.06	0.06	177.85	<0.0001	
Location	1	0.17	0.17	521.98	<0.0001	
Year*Location	1	0.03	0.03	109.99	<0.0001	
Genotype	123	0.54	0.004	13.4	<0.0001	
Year*Genotype	123	0.09	0.0007	2.25	<0.0001	
Location*Genotype	123	0.04	0.0003	0.96	0.59	
Year*Location*Genotype	123	0.04	0.0003	0.99	0.51	
Error	466	0.15	0.0003			

Table 3. Summary of SSR markers used in screening parents and polymorphic markers used in screening the 120 RIL population derived from the cross KS4303sp x PI 407818 B.

Chromosome	No. SSR markers screened	Coverage (cM)	No. polymorphic markers	Coverage (cM)
1	25	102.96	9	68.27
2	27	137.06	9	80.25
3	21	76.21	4	65.39
4	18	127.77	0	0
5	25	101.57	7	83.39
6	25	125.08	5	110.24
7	24	112.79	9	82.79
8	35	136.13	8	126.21
9	23	115.22	2	45.71
10	20	124.36	4	62.48
11	20	91.62	3	67.84
12	20	100.86	2	17.99
13	30	140.43	8	117.27
14	23	107.56	5	40.93
15	22	68.01	2	23.65
16	19	78.59	4	30.12
17	22	104.21	3	48.38
18	40	108.7	9	95.47
19	25	115.07	5	91.35
20	22	82.28	7	59.94
Total	**486**	**2156**	**105**	**1318**
Average	**24**	**107.82**	**5.25**	**12.50**

Table 4. Summary of SSR markers used in screening parents and polymorphic markers used in screening the 124 RIL population derived from the cross PI 408052 C x PI 407818 B.

Chromosome	No. SSR markers screened	Coverage (cM)	No. polymorphic markers	Coverage (cM)
1	25	102.96	9	92.96
2	27	137.06	9	80.25
3	21	76.21	4	61.25
4	18	127.77	2	26.74
5	25	101.57	6	70.1
6	25	125.08	5	101.02
7	24	112.79	2	8.73
8	35	136.13	4	71.44
9	23	115.22	2	9.7
10	20	124.36	3	62.48
11	20	91.62	3	58.2
12	20	100.86	5	81.61
13	30	140.43	7	98.13
14	23	107.56	8	56.85
15	22	68.01	2	12.27
16	19	78.59	6	43.79
17	22	104.21	4	75.9
18	40	108.7	9	98.16
19	25	115.07	6	91.35
20	22	82.28	5	55
Total	**486**	**2156**	**101**	**1255**
Average	**24**	**107.82**	**5.05**	**12.43**

Table 5. Single marker analysis of variance for seed calcium content of 120 RIL from cross KS4303sp x PI 407818 B grown in Fayetteville, AR (FAY) and Keiser, AR (KEI) in 2008 and 2009.

		2008		2009		Combined†	
Chromosome	**Position (cM)**	**FAY**	**KEI**	**FAY**	**KEI**		**R^2**
1	17.52	Satt184*	Satt184*	Satt184 ****	Satt184**	Satt184****	0.15
1	36.23		Sat_353*	Sat_353*	Sat_353**	Sat_353**	0.10
1	48.14			Satt342**	Satt342 ****	Satt342***	0.13
1	56.43	Satt254*		Satt254**	Satt254 ****	Satt254****	0.14
1	108.89			Satt147*	Satt147*	Satt147*	0.08
2	25.6	Satt095*	Satt095*	Satt095****	Satt095 ****	Satt095****	0.14
2	72.89			Satt141*	Satt141*	Satt141*	0.05
2	75.41		Satt600*	Satt600*		Satt600*	0.05
2	84.04			Satt041*	Satt041**	Satt041*	0.07
7	1.00					**Sat_391‡**	
7	66.99	Satt175**	Satt175**	Satt175 ****	Satt175 ****	Satt175****	0.34
7	75.57	Satt677*		Satt677*	Satt677**	**Satt677**‡**	0.10
7	76.41	Sat_288*		Sat_288*	Sat_288*	Sat_288*	0.08
8	84.09		Sat_199*	Sat_199*	Sat_199*	Sat_199**	0.11
8	90.84		Satt377*	Satt377**	Satt377**	Satt377**	0.13
8	116.64	Sat_377*	Sat_377*	Sat_377*	Sat_377*	**Sat_377*‡**	0.10
8	154.10					**Satt220‡**	
9	32.96			Satt055**	Satt055*	Satt055*	0.09
10	86.86	Satt123*	Satt123*	Satt123 ****	Satt123*	Satt123****	0.16
10	118.14			satt153**		Satt153*	0.07
13	16.08	Satt252*		Satt252*	Satt252*	Satt252**	0.10
13	135.12			Satt656*	Satt656*	Satt656*	0.07
16	37.04					Satt414*	0.05

16	67.79	Satt547*	Satt547**	Satt547 ****	Satt547*	Satt547****	0.14
17	80.19			Satt461*	Satt461*	Satt461*	0.90
17	105.45					Satt186*	0.07
18	27.48		Sat_315*	Sat_315*		Sat_315*	0.08
18	52.94	Satt594*		Satt594*		Satt594*	0.06
18	87.2	Satt564*		Satt564*		Satt564*	0.07
20	36.59	Sat_174*	Sat_174*	Sat_174**	Sat_174**	**Sat_174*****‡**	0.12
20	46.20					**Satt354‡**	
20	75				Sat_170*	Sat_170*	0.05
20	82.78			Satt292*		Satt292*	0.07

† refers to the pooled data from two locations in two years.

‡ refers to a previously reported marker for calcium content by Zhang et al. (2009).

45 *, **, ***, and **** represent significant association with calcium content at P ≤0.05, 0.01, 0.001, and 0.0001 respectively.

Table 6. Single marker analysis of variance for seed calcium content of 124 RIL from the cross PI 408052 C x PI 407818 B grown in Fayetteville, AR (FAY) and Keiser, AR (KEI) in 2008 and 2009.

		2008		2009		Combined†	
Chromosome	**Position (cM)**	**FAY**	**KEI**	**FAY**	**KEI**		**R^2**
1	48.14	Satt342*		Satt342*		Satt342*	0.05
1	77.49	Satt077*	Satt077*	Satt077*	Satt077**	Satt077***	0.07
1	106.69			Sat_408*			
5	27.66	Satt042*		Satt042*	Satt042*	Satt042*	0.05
5	42.8	Sat_356***	Sat_356***	Sat_356***	Sat_356***	Sat_356****	0.12
7	1.00					**Sat_391‡**	
7	33.47			Satt567**	Satt567**	Satt567**	0.07
7	75.57					**Satt677‡**	
8	116.64	Sat_377**	Sat_377*	Sat_377**	Sat_377**	**Sat_377**‡**	0.06
8	116.73	Satt470**		Satt470*		Satt470*	0.06
8	118.64	Sat_040*			Sat_040**	Sat_040*	0.07
8	154.10					**Satt220‡**	
10	24.61	Sat_318*			Sat_318*	Sat_318*	0.05
10	119.5	Satt243*	Satt243*			Satt243*	0.07
12	53.35	Satt541*	Satt541*	Satt541**	Satt541**	Satt541**	0.08
17	87.16	Sat_338*			Sat_338*	Sat_338*	0.05
18	18.25	Satt217*	Satt217*	Satt217*	Satt217**	Satt217**	0.08
18	29.03	Sat_290*	Sat_290***	Sat_290**	Sat_290**	Sat_290**	0.10
18	43.78	Satt115***	Satt115**	Satt115**	Satt115**	Satt115**	0.13
18	45.49	Sat_358*	Sat_358*	Sat_358**	Sat_358**	Sat_358**	0.08

† refers to the pooled data from two locations in two years.

‡ refers to previously reported markers for calcium content by Zhang et al. (2009).

*, **, ***, and **** represent significant association with calcium content at P ≤0.05, 0.01, 0.001, and 0.0001, respectively.

Table 7. Mean allelic effect of SSR marker alleles on seed calcium content in 120 RIL population from the cross KS43030sp x PI 407818 B grown in Fayetteville, AR and Keiser, AR in 2008 and 2009.

	2008			2009			Combined			
SSR marker/ Chromosome	**PI 407818 B**	**KS4303sp**	**Allelic difference**	**PI 407818 B**	**KS4303sp**	**Allelic difference**	**PI 407818 B**	**KS4303sp**	**Allelic difference**	**R^2**
Satt184 (1)	0.25	0.24	0.01	0.26	0.24	0.02	0.25	0.24	0.01	0.15
Satt095 (2)	0.25	0.24	0.01	0.26	0.24	0.02	0.25	0.24	0.01	0.14
Satt175 (7)	0.25	0.24	0.02	0.26	0.24	0.01	0.26	0.24	0.02	0.34
Sat_377 (8)	0.25	0.23	0.02	0.26	0.24	0.02	0.25	0.24	0.02	0.10
Satt123 (10)	0.25	0.23	0.02	0.27	0.25	0.02	0.26	0.24	0.02	0.16
Satt547 (16)	0.25	0.24	0.01	0.27	0.26	0.01	0.26	0.25	0.01	0.14
Sat_174 (20)	0.24	0.23	0.01	0.27	0.25	0.02	0.25	0.24	0.02	0.12

Table 8. Mean allelic effect of SSR marker alleles on seed calcium content in 124 RIL population from the cross PI 408052 C x PI 407818 B grown in Fayetteville, AR and Keiser, AR in 2008 and 2009.

	2008			2009			Combined			
SSR marker/ Chromosome	**PI 407818 B**	**PI 408052 C**	**Allelic difference**	**PI 407818 B**	**PI 408052 C**	**Allelic difference**	**PI 407818 B**	**PI 408052 C**	**Allelic difference**	**R^2**
Satt077 (1)	0.27	0.25	0.02	0.27	0.25	0.02	0.27	0.25	0.02	0.07
Sat_356 (5)	0.26	0.25	0.01	0.26	0.25	0.01	0.26	0.25	0.01	0.12
Sat_377 (8)	0.27	0.25	0.02	0.26	0.25	0.02	0.26	0.25	0.02	0.06
Satt541 (12)	0.27	0.26	0.01	0.27	0.25	0.02	0.27	0.25	0.02	0.08
Satt217 (18)	0.26	0.25	0.02	0.27	0.25	0.02	0.26	0.25	0.02	0.08
Sat_290 (18)	0.26	0.25	0.02	0.26	0.24	0.02	0.26	0.24	0.02	0.10
Satt115 (18)	0.27	0.25	0.02	0.27	0.25	0.02	0.27	0.25	0.02	0.13
Sat_358 (18)	0.26	0.25	0.02	0.26	0.25	0.02	0.26	0.25	0.02	0.08

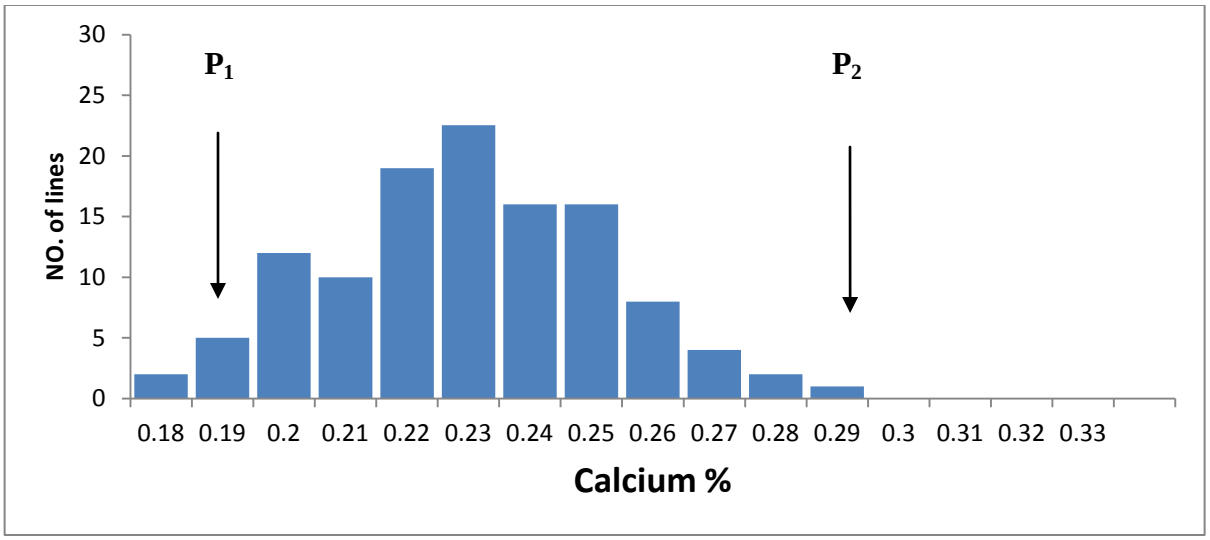


Figure 1. Frequency distribution of calcium content of 120 RIL from the cross KS4303sp (P1) x PI 407818 B (P2), grown in Fayetteville, AR in 2008.

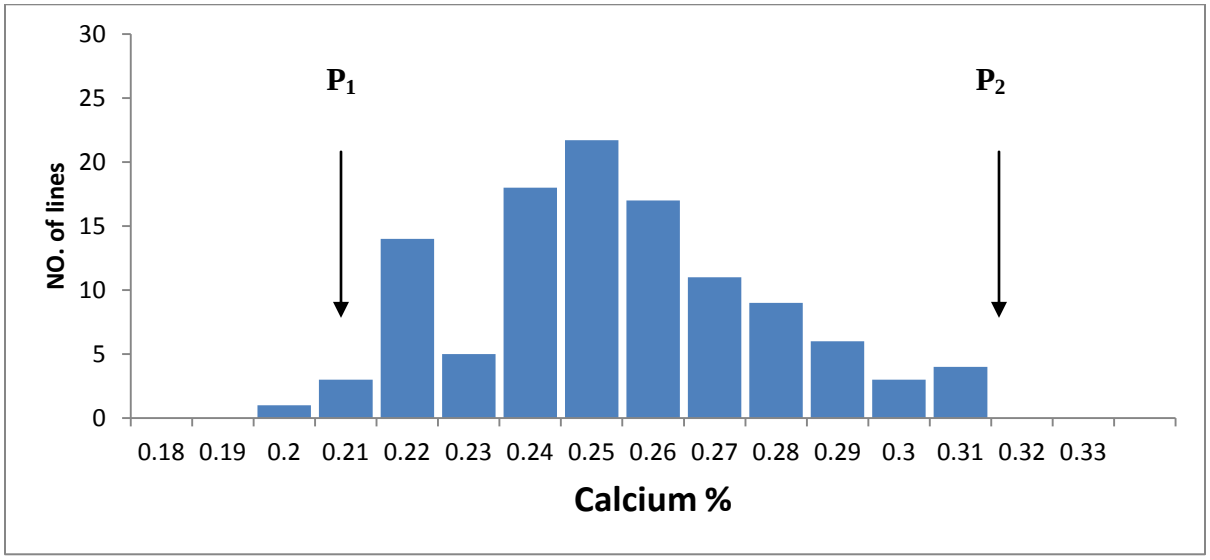


Figure 2. Frequency distribution of calcium content of 120 RIL from the cross KS4303sp (P1) x PI 407818 B (P2), grown in Keiser, AR in 2008

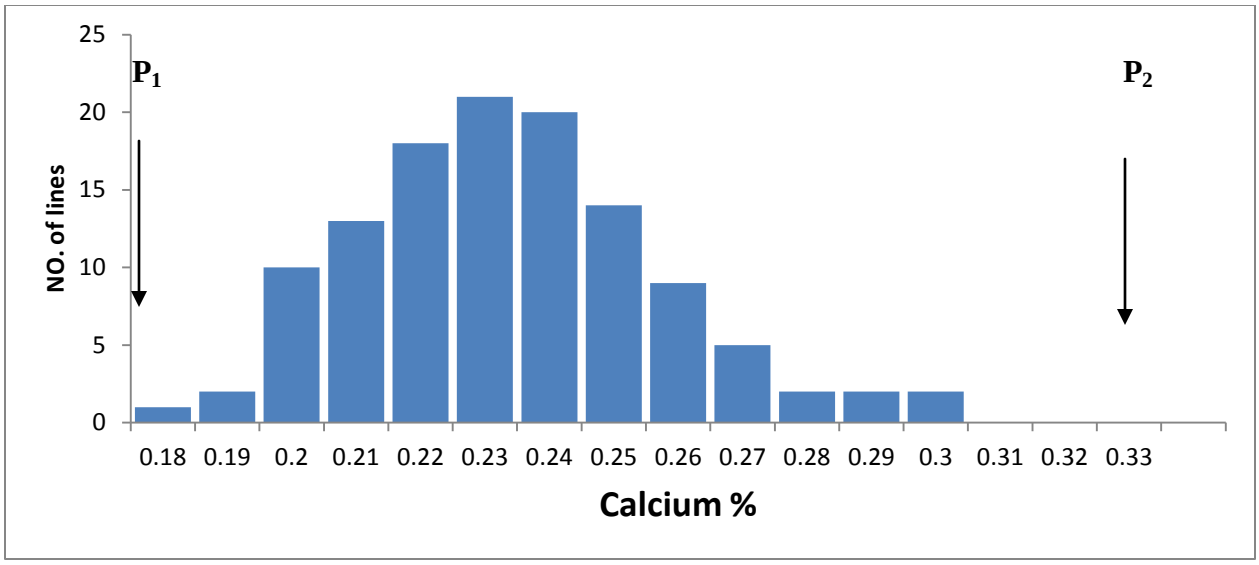


Figure 3. Frequency distribution of calcium content of 120 RIL from the cross KS4303sp (P1) x PI 407818 B (P2), grown in Fayetteville, AR in 2009

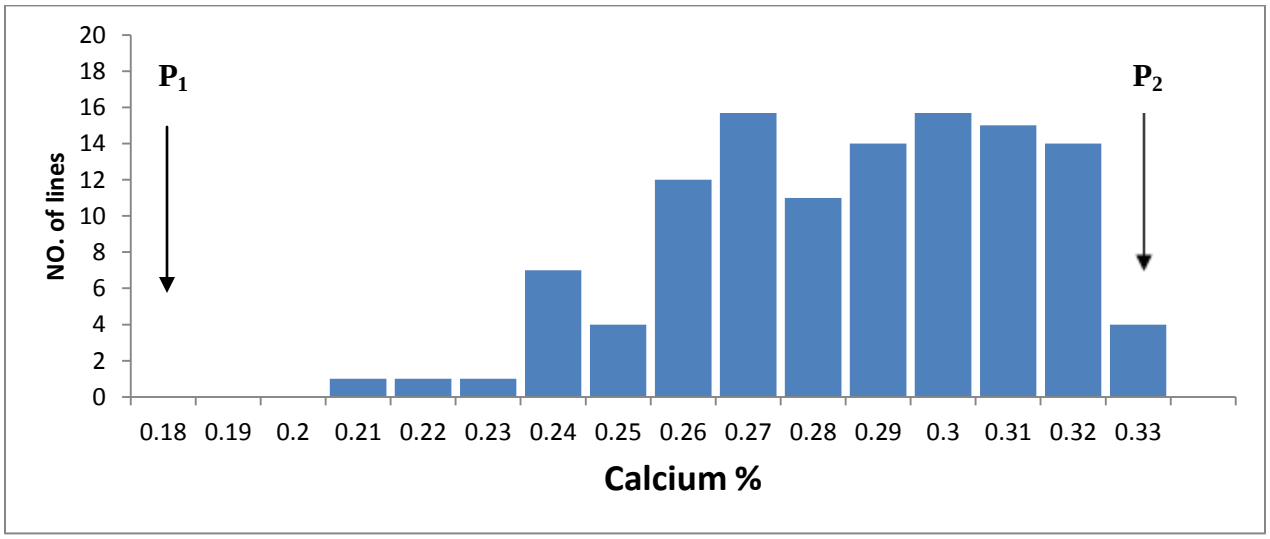


Figure 4. Frequency distribution of calcium content of 120 RIL from the cross KS4303sp (P1) x PI 407818 B (P2), grown in Keiser, AR in 2009

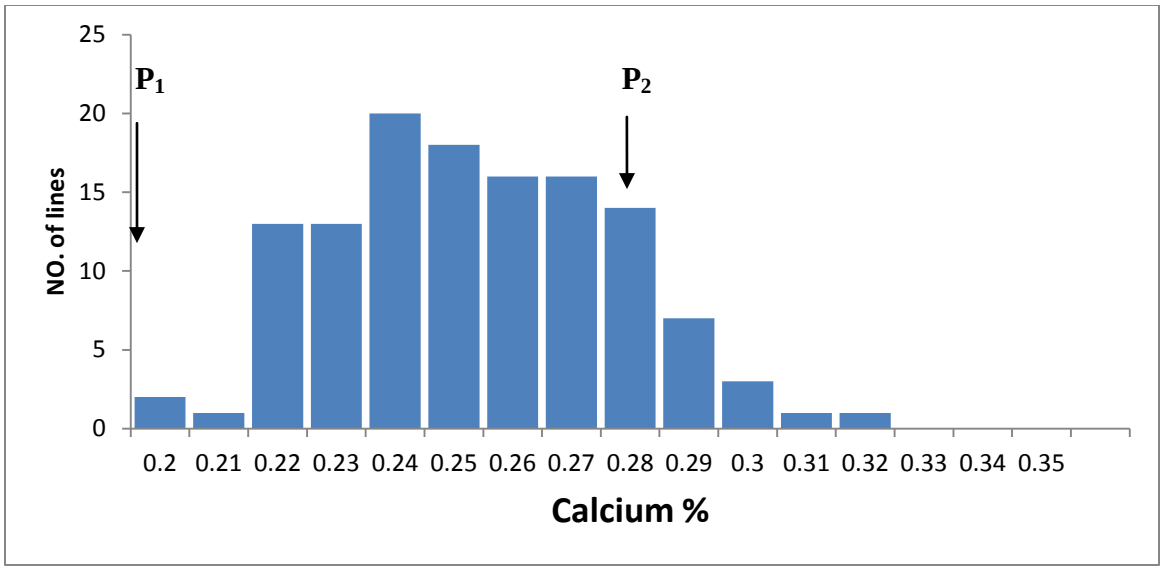


Figure 5. Frequency distribution of calcium content of 124 RIL from the cross PI 408052 C (P1) x PI 407818 B (P2), grown in Fayetteville, AR in 2008

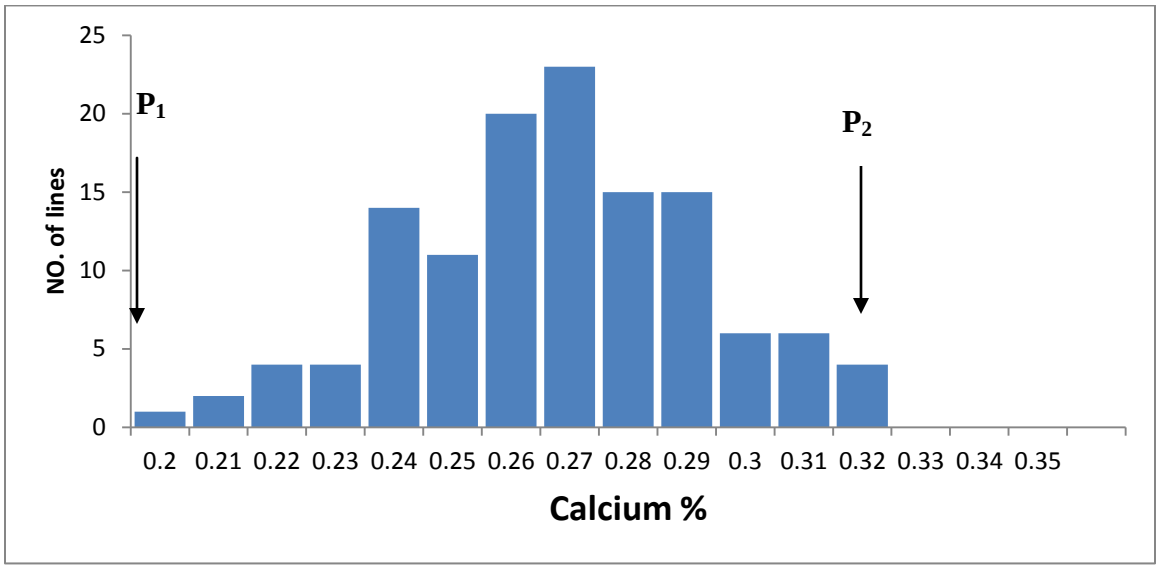


Figure 6. Frequency distribution of calcium content of 124 RIL from the cross PI 408052 C (P1) x PI 407818 B (P2), grown in Keiser, AR in 2008.

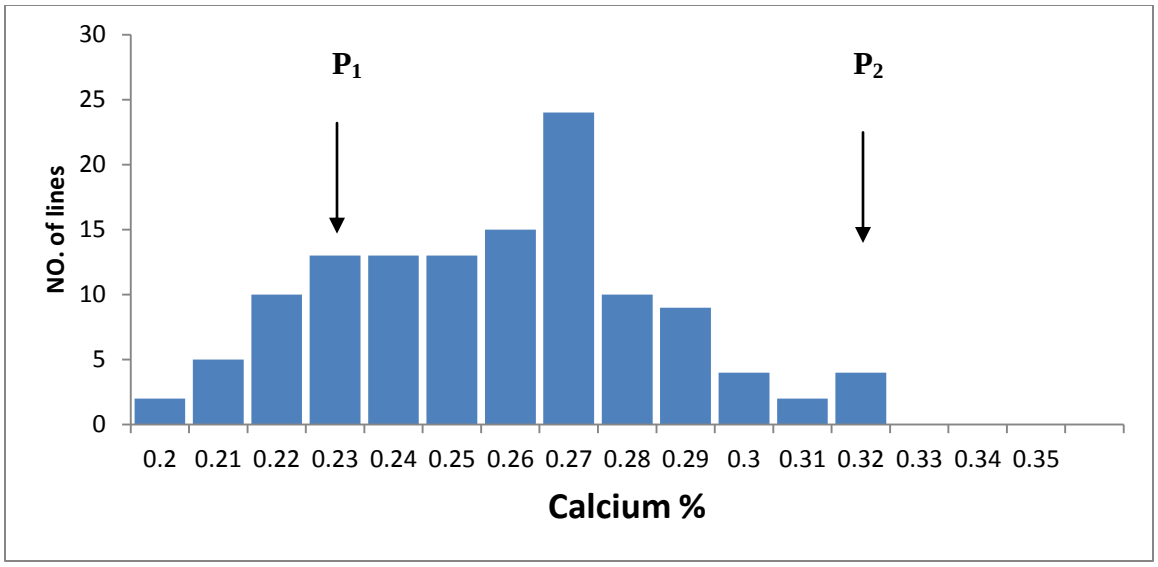


Figure 7. Frequency distribution of calcium content of 124 RIL from the cross PI 405082 C (P1) x PI 407818 B (P2), grown in Fayetteville, AR in 2009.

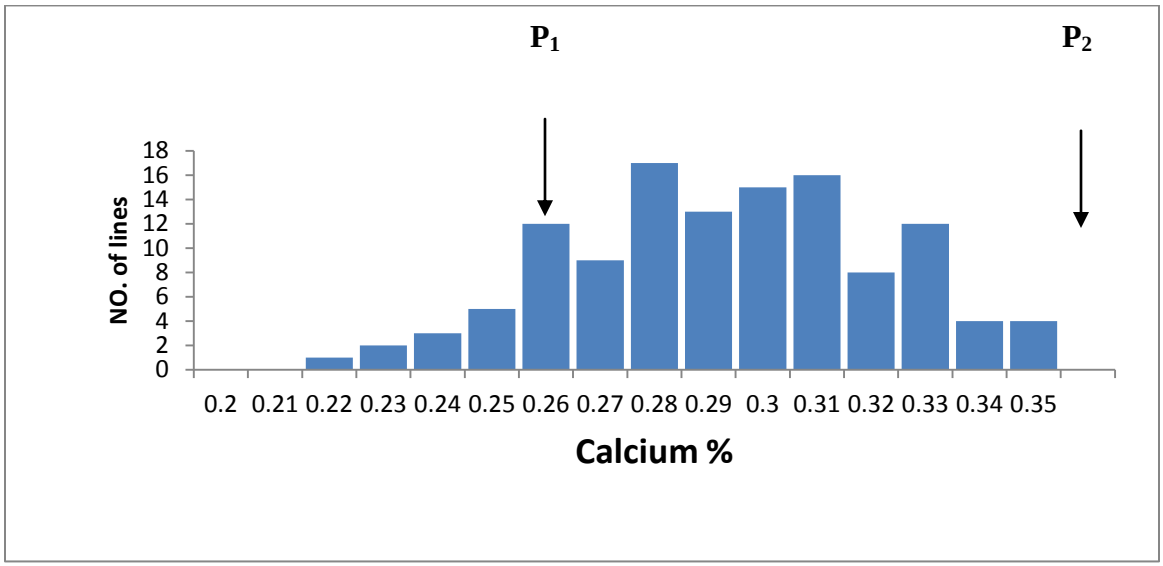


Figure 8. Frequency distribution of calcium content of 124 RIL from the cross PI 405082 C (P1) x PI 407818 B (P2), grown in Keiser, AR in 2009.

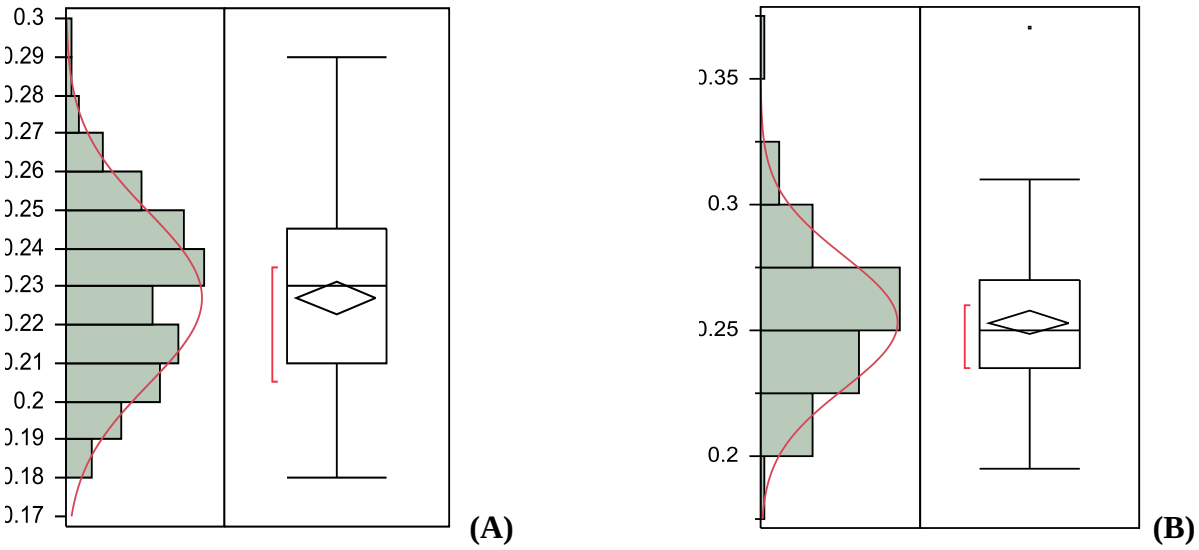


Figure 9. Normal distribution of seed calcium content of 120 RIL from the cross KS4303sp x PI 407818 B, grown in Fayetteville, AR (A) and in Keiser, AR (B) in 2008.

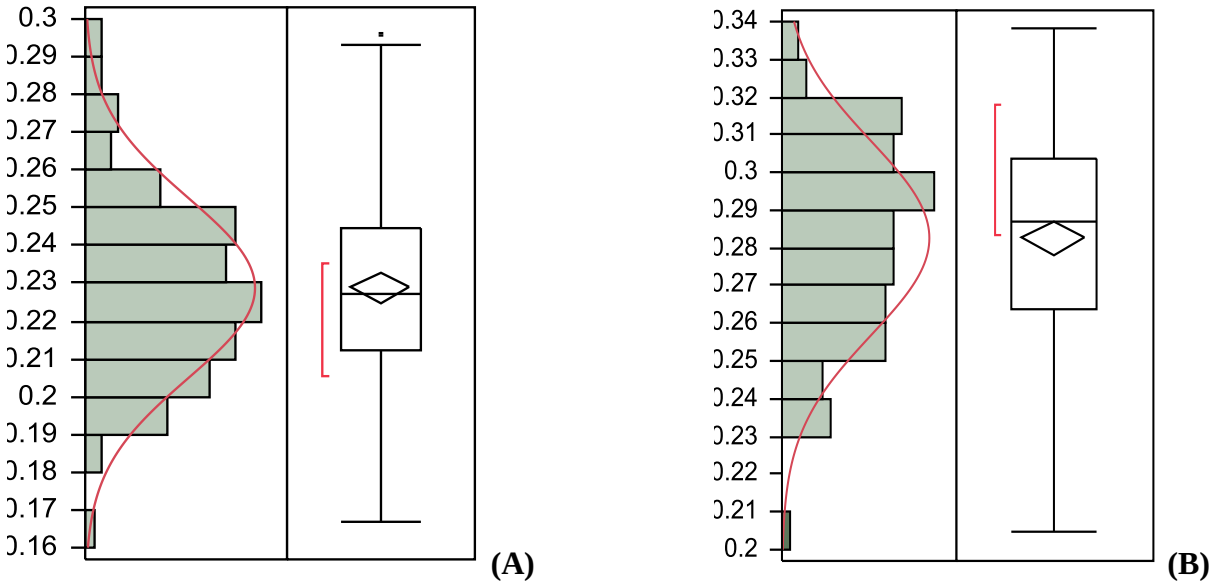


Figure 10. Normal distribution of seed calcium content of 120 RIL from the cross KS4303sp x PI 407818 B, grown in Fayetteville, AR (A) and in Keiser, AR (B) in 2009.

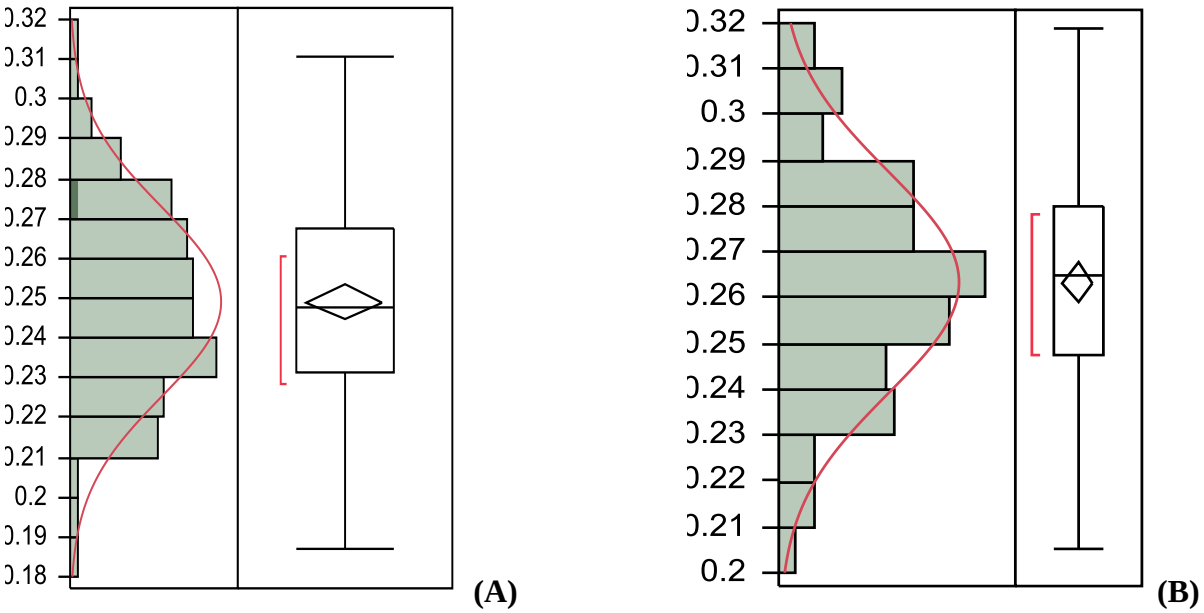


Figure 11. Normal distributions of seed calcium content of 124 RIL from the cross PI 408052 C x PI 407818 B, grown in Fayetteville, AR (A) and in Keiser, AR (B) in 2008.

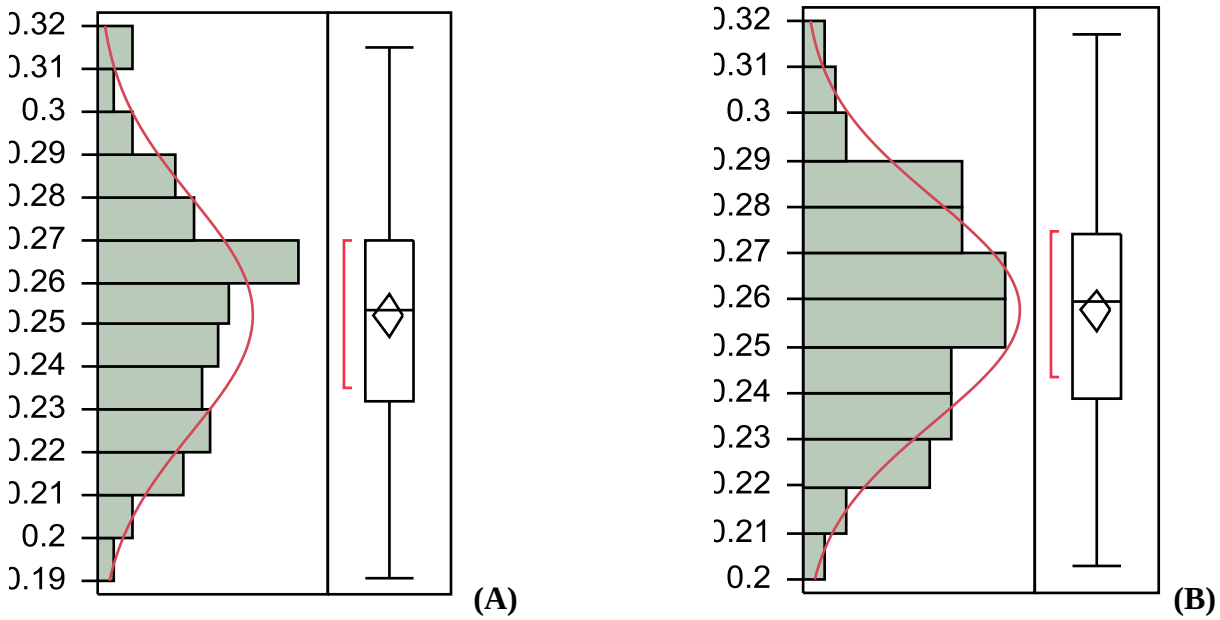


Figure 12. Normal distributions of seed calcium content of 124 RIL from the cross PI 408052 C x PI 407818 B, grown in Fayetteville, AR (A) and in Keiser, AR (B) in 2009.

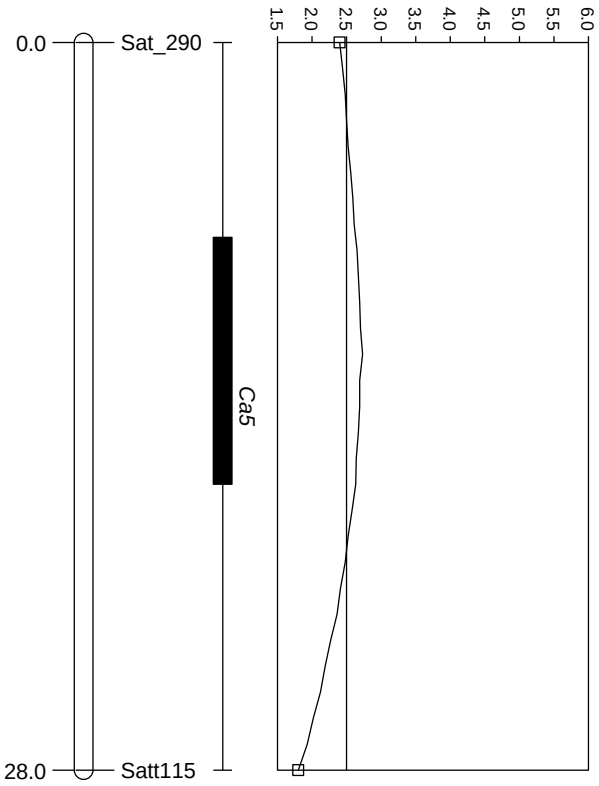


Figure 13. Interval mapping of QTL on chromosome 18 for soybean seed calcium content of 124 RIL derived from the cross PI 408052 C x PI 407818 B using combined data across locations (Fayetteville and Keiser) and the years (2008 and 2009).

Chapter III

Putative Quantitative Trait Loci (QTL) Associated with Seed Hardness of Soybean

ABSTRACT

Soybean seed hardness is a determining factor in texture and is therefore an important attribute for food grade soybean varieties. Cooked seed hardness is particularly important for soyfood types such as natto and edamame where cooking is required. Texture analysis is time consuming and labor intensive. Molecular markers and QTL for seed hardness will facilitate the development of elite cultivars with soft seeds through marker assisted selection (MAS). The objectives of this study were to identify new QTL associated with seed hardness and confirm previously reported QTL for this trait. A population of 120 RIL population derived from KS4303sp (soft seeded) x PI 407818 B (hard seeded) were screened with 486 SSR markers and tested for cooked seed hardness. The $F_{2:4}$ and $F_{2:5}$ lines were grown in a randomized complete block design (RSBD) with two replications, at Fayetteville and Keiser, AR in 2008 and 2009. Interval mapping revealed a new putative QTL for seed hardness (*Ha3*) residing in an approximate 46 cM region and linked to markers Satt547 and Satt414 on chromosome 16. This QTL contributed 32% of seed hardness in the KS4303sp x PI 407818 B population. Single marker analysis for the data over years and across locations revealed 12 markers on seven chromosomes associated with seed hardness. Briefly, Satt254 on chromosome 1, Satt677 on chromosome 7, and Satt292 on chromosome 20 were significantly associated with seed hardness in three out of four environments. A chromosomal region previously reported for hardness (QTL *Ha2*) contained a marker (Satt254 on chromosome 1) was stable across environments for seed hardness in current study. These identified QTL/markers and confirmed QTL (*Ha2*) region in current study will facilitate MAS for seed hardness in soybean breeding programs for food-grade soybean.

INTRODUCTION

Soybean (*Glycine max* (L.) Merrill) is worldwide important leguminous with high contents of protein, oil, and other physiologically active substances. This nitrogen fixing legume that is commonly grown as a source of vegetable oil and protein rich animal feed is also known for soyfood production due to their nutritional qualities. Soyfood has been consumed in Asia for centuries and it is becoming popular in western diet for its health benefits and nutritional value. Soyfood is categorized as whole or cracked, and fermented or non-fermented. Most common products that are consumed in the US include green vegetable soybean, known as "edamame", soy sprouts, soymilk, and tofu known as soybean curd. These foods require specific characteristics of seed such as large or small size, high protein, high sugar, and light hilum color. Another important requirement for soyfood production is texture which can be affected by many factors such as seed size, seed coat water permeability (Qutob et al., 2008), imbibition, dormancy, seed coat structure, and chemical compositions (Shao et al., 2007). Cooked seed hardness is critical for soyfood types that require processing and cooking such as edamame and natto.

For crop plants seed hardness is not only a biological problem, it is also an economical issue. Although seed hardness provides benefits for resistance to seed spoilage when storing conditions are not optimim and resistance to seed coat pathogens, it also causes impermeability where seed does not absorb enough water and it is known as "stone seed." Less permeable seeds require scarification which is an extra step (Argel and Paton, 1999). Stone seeds cause a major issue on soyfood production because uniform water absorbability is required during processing; at the end, it affects the texture of soyfood. Soybean lines with stone seeds are less desired or rejected by soyfood manufacturers especially when it is for whole bean products. Stress conditions tend to cause harder seeds. In addition, higher calcium content can be the cause of

stone seeds as well. Zhang et al., (2008) indicated that small seeds tend to have higher calcium content, consequently, cause stone seeds or harder texture. Their results showed that larger seeds had lower calcium content and were softer when they were cooked (Zhang et al., 2008, Chen at al., 2001). Taira (1990) also showed that seed size is negatively correlated with seed water absorption and cooked seed hardness. This negative correlation between seed size and hardness was also observed in other species (Yaklich et al., 1986, Ragus 1987).

Soybean seed coat directly affects water permeability; the permeability is dependent on the thickness of the seed coat (Werker ey al., 1979). Low pectin content and high hemicelluloses can reduce hydrophlicity which may lead to hard seeds (Mullin et al., 2001). Inheritance studies showed that seed impermeability and pigmentation loci have positive association (Keim et al., 1990). Black coated soybean seeds have lower imbibition rate (Kuo, 1989). Qutob et al., (2008) suggested possible pleiotropic effect between pigmentation loci and permeability through their activity towards polyphenolic substrates that become impregnanted in the cell wall. More domesticated genotypes have less hard seeds than wild or primitive cultivated types due to their less thick seed coat (Lush and Evans, 1980). These previous research suggest the role of seed coat in water intake and role of water intake in seed hardness. Ma et al., (2004) compared six different soybean varieties and indicated the role of small cracks in the cuticle. They concluded that seeds with no cracks had high impermeability. Cracks were very microscopic and not visible to naked eye.

High temperature, low air relative humidity, and water stress cause harder seeds (Argel and Paton, 1999). Depending on weather conditions, amount of hard seed amount vary according to year and geographic location (Calero et al., 1981). Genetic factors, geographic location, and

weather conditions during plant growth determine the amount of hard seeds, while seed permeability is determined by genotype (Tinius, 1991).

Keim et al., (1990) found five restriction fragment length polymorphism (RFLP) markers on chromosome 15 for seed hardness in a segregating population from a cross between a wild and a cultivated soybean. Watabane et al., (2004) identified three seed hardness QTLs (*RAS1-3*) on chromosomes 2, 6, and 20, respectively using the information based on percentage of seeds that absorbed water. However, genetic basis of hard seedness has not been clearly understood. Additional information is needed to understand the genetics of the seed hardness.

Identification of QTL for seed hardness was primarily focused on permeability in research conducted by Watanabe et al., (2004) and Sakamoto et al., (2004). For food-grade soybeans, however, cooked seed hardness is important, even though seed may not have problems with permeability and imbibition. Using SSR markers, Zhang et al., (2008) identified two QTL for cooked seed hardness on chromosomes 1 and 19. Soybean seed hardness and its role in texture are determining factors in soyfood production. Information on genetic control of seed hardness will be a useful in selecting soybean lines with desirable seed hardness which will facilitate the soybean breeding process for specialty soybeans for soyfood market.

The objectives of this study were to identify SSR markers and QTL associated with seed hardness of soybean and confirm previously reported QTL.

MATERIALS AND METHODS

Population Development and Field Experiment

A hard seeded soybean line KS4303sp was crossed with a soft seeded soybean line PI 407818 B. The cross was made in spring of 2007 in the greenhouse. The F_1 plants were space planted in the field at the Arkansas Research and Extension Center affiliated with the University

of Arkansas in Fayetteville, AR in summer 2007. Flower color and leaf shape were used as morphological marker to identify true hybrids. F_2 plants were space planted and harvested individually in a winter nursery in Costa Rica in winter of 2007 and 100 seeds from each F_2 plant was taken to obtain next generation, $F_{2:3}$ lines. Each $F_{2:3}$ line was grown in a 3 m row and each row was bulk harvested in the Costa Rica winter nursery in spring 2008. A total of 200 $F_{2:3}$ lines were developed from cross KS4303sp X PI 407818 B. One hundred twenty progeny lines from the cross KS4303sp X PI 407818 B with adequate amount of seeds were selected as a source of population for QTL mapping. In summer 2008, $F_{2:4}$ lines were grown in the field with complete randomized block design (RCBD) with two replications at two locations Fayetteville, AR and Keiser, AR. Parental lines KS4303sp and PI 407818 B were included as checks in the test. Each line was grown in a single row plot with 3 m in length and 0.95 m row spacing. Irrigation was conducted according to University of Arkansas irrigation schedule. Once plants in a row reached 95% maturity, they were bulk harvested either with a combine or a plant thresher. Field plots were managed using standard cultural practices adopted for full season soybean production in Arkansas. In 2009, $F_{2:5}$ lines were grown in RCBD with two replications at two locations; Fayetteville, AR and Keiser, AR. Plots were managed in the same fashion as described for $F_{2:4}$ lines.

DNA extraction

Leaf tissues were collected from the uppermost partly developed trifoliate leaf of every plant of each $F_{2:4}$ line. The bulked tissue samples were ground with a mortar and pestle in liquid nitrogen and stored in a -20^{o}C freezer. Total genomic DNA was isolated using the CTAB (hexadercyltrimethyl ammonium bromide) method (Kisha et al., 1997). Briefly, 0.50 mg powder obtained from ground leaves were transferred to a 2.0 ml microtube and 750 µl of CTAB

extraction buffer was added followed by incubation at 65^{o}C for 60 min with inverting every 10 min. Tubes were cooled and 500 µl chloroform:isoamyl alcohol (24:1) was added and inverted several times followed by centrifuging samples at 13,500 for 15 min. The aqueous layer was transferred to a 2.0 ml microtube and centrifuged for 5 min. which lead to DNA precipitation. DNA was washed with 95% ethanol solution. Supernatant was poured out and DNA pellets were allowed to air dry at room temperature overnight. 250 µl of 0.1 x TE was added over dried DNA and left in temperature for a day to dissolve DNA. Each sample was checked for its concentration using Bio-Tek PowerWave XS Microplate Spectrophotometer.

SSR polymorphism analysis

Polymerase chain reaction (PCR) and simple sequence repeat (SSR) primer genotyping were performed in a 96-well or 384-well BIO-RAD iCycler thermocycler (Bio-Rad, Hercules, CA). PCR reaction was performed using 4 µl of 20ng/µl template DNA, 1.0 µM of 0.5 µM forward and reverse primers, 0.2 µl of Go Taq Flexi polymerase, 0.9 µl of 2.5 mM dNTP (nucleotides) mixture, 1.8 µl of 2.5 mM $MgCl_2$, and 4.3 deionized H_2O. PCR reaction were at 94^{o}C for 4 min. followed by 33 cycles at 94^{o}C for 25 s for denaturation, 47^{o}C for 25 s for annealing, 68^{o}C for 25 s for extension, and 72^{o}C for 5 min. for final extension after the last cycle. Amplified PCR products were separated on a 6% non-denaturing polyacrylamide gel on a Megagel electrophoresis system (C.B.S. Scientific) and stained with ethidium bromide (Wang et al., 2003). The gels were visualized under UV fluorescence and digital images were taken to score individuals for marker alleles. Parental genotypes, KS4303sp and PI 407818 B, were screened using 486 SSR markers from 20 chromosomes of soybean. Markers scored as monomorphic when both parents exhibited amplification products of the same size and as polymorphic when parents had fragments in different size. Only polymorphic markers were

chosen to screen the progenies. There were 105 polymorphic markers for the cross KS4303sp x PI 407818 B. Primers for the SSR markers used in this study, their sequences, and their integration in soybean molecular linkage map are publicly available online at http://129.186.26.94/ssr.html (Cregan et al., 1999, Song et al., 2004).

Hardness test

Thirty grams of unbroken uniform seeds from each progeny line was weighed and soaked in heat resistant plastic boxes containing 150 mL water at ambient temperature for 16 hours. Seeds were recovered from soaking water with a sieve and blot dried on paper towels. Stone seeds that did not absorb water during soaking process were removed from soaked seeds. Soaked seed samples were pressure cooked in heat resistant plastic boxes at 121.1^{o}C and 1.2 kg cm^{-2} for 20 min. Hardness test of 30 g cooked seeds from each sample were conducted in two replications using a TMS Texture System (TMS-2000, Food Technology Corp., Sterling, VA) equipped with a 16-blade shear cell. The maximum force to compress cooked seeds in newtons (N) was determined as seed hardness (Song et al., 2003).

Linkage map construction and QTL analysis

Association between molecular data and phenotypic data used to initially identify candidate QTL was analyzed by single-factor analysis of variance with the PROC GLM procedures of SAS software v9.2 (SAS Institute, 2002). Linkage mapping and QTL analysis procedures were performed by using JoinMap 3.0 (Van Ooijen and Voopris, 2001) with Kosambi function (Kosambi, 1944) and WinQTL cartographer (Basten et al., 1999) with single marker analysis (SMA) and interval mapping (IM) to identify QTL associated with calcium content. For SMA, $p<0.05$ was used as a threshold for significant markers (Cornelious et al., 2005). In IM, the empirical significance threshold was determined by 1000 permutations, with a

walk speed of 1 cM, and a significance level of alpha =0.05 (Cornelious et al., 2005). Mapchart (Voopris, 2002) was used to create the LOD plots according to the data from WinQTL Cartographer.

Statistical data analysis

Phenotypic data for seed hardness collected in 2008 and 2009 were analyzed both separately and as combined using the PROC MEANS procedure in SAS software v9.2 (SAS Institute, 2002). The PROC GLM procedure was used to determine variance components for seed hardness with location and replication as random blocking factors. Distribution of seed hardness and normality test at two locations and in two years were analyzed using JMP software.

RESULTS AND DISCUSSION

Table 1 presents the analysis of variance for cooked seed hardness of 120 RIL from the cross KS4303sp x PI 407818 B grown at two locations over two years. The model was significant ($p = <0.0001$ and $R^2 = 0.70$) and accounted for differences for the following sources of variation: year, location, year x location, genotype, genotype x year, genotype x location, and year x location x genotype. Effects of year, genotype, year x location, and year x genotype on seed hardness were statistically significant. Location and genotype x location effects were not significant, allowing for data to be combined across locations when necessary. In addition, the year x location x genotype effect was not statistically significant, which allowed pooled data to be used in the over-all analysis. ANOVA of the 120 RIL population also showed that year x location was the most important source of variation for seed hardness followed by year and then by genotype. Therefore, it is logical to assume that each year and location combination represents a specific environment. It is expected that each location would have different soil type and different temperature and rainfalls in different years. Therefore, in practical breeding, lines

need to be evaluated in multiple environments across locations and over years for an effective selection for seed hardness.

Figures 1 to 4 show the frequency distribution of cooked seed hardness of 120 RIL derived from the cross KS4303sp x PI 407818 B at two locations in two years. A normality test of these data showed normal distribution of cooked seed hardness in both years and locations (Fig. 5 and 6). In 2008, frequency distribution of seed hardness was in a similar range in both locations with a population mean slightly more towards to the hard side in Fayetteville. Cooked seed hardness ranged from 260 to 590 N with a mean of 394 N at Fayetteville and from 200 to 560 N with a mean of 344 N at Keiser in 2008. In 2009, seed hardness ranged from 230 to 470 N with a mean of 328 N at Fayetteville and from 230 to 530 N with a mean of 378 N at Keiser. Cooked seed hardness of soft-seeded parent KS4303sp ranged from 256 to 296 N with an average of 277 N and hard-seeded parent PI 407818 B from 420 to 516 N with a mean of 465 N. These phenotypic data was analyzed separately for each location and each year prior to performing the combined data analysis.

Out of a total of 486 markers screened, 105 SSR markers were polymorphic between the parental lines KS4303sp and PI 407818 B (Table 2). These polymorphic markers were distributed in 19 out of 20 chromosomes of soybean and they were used to screen the 120 RIL population. A larger numbers of polymorphic markers were found in chromosomes 1, 2, 7, 8, 13, and 18 than other chromosomes. On average, the density of polymorphic markers ranged from two to nine per chromosome with an average of a marker per each 12.5 cM. Prior to linkage analysis, each polymorphic marker was analyzed for the goodness-of-fit to the expected allelic segregation ratio (1:2:1) in the population.

Single marker analysis of the 120 RIL in each year and at each location revealed one marker each on chromosomes 1 and 7, 3 markers on chromosome 16, and one marker on chromosome 20 at Fayetteville in 2008 that are associated with seed hardness; while at Keiser in 2008, two markers on chromosome 16 and one marker on chromosome 18 were identified to be associated with seed hardness (Table 3). In 2009 at both locations, a large number of markers were identified to be associated with seed hardness: one on chromosome 1, three on chromosome 7, one on chromosome 11, two on chromosome 13, three on chromosome 16, two on chromosome 18, and two on chromosome 20; and at Keiser, one marker on chromosome 1, three on chromosome 7, one on chromosome 11, two on chromosome 13, two on chromosome 16, and one on chromosome 20 (Table 3). Combined data analysis revealed a total of 12 markers on seven different chromosomes that are associated with seed hardness (Table 3). Some of the linked markers were stable over years and locations except for Keiser in 2008 where there were some missing data due to poor germination of seeds coming back from the winter nursery.

SSR markers Satt254 on chromosome 1, Satt677 on chromosome 7, Satt414, Satt406, and Satt547 on chromosome 16, Satt292 on chromosome 20 were consistent and stable in 3 out of the 4 environments. Markers Satt414, Satt406, and Satt547 on chromosome 16 explained 11, 12, and 15 % of phenotypic variation with a p value of 0.0001 and 0.001, and 0.001, respectively. The rest of the significant markers explained phenotypic variation in a range of 5-8% (Table 3). Interval mapping using WinQTL cartographer for the SSR markers revealed a new QTL (*Ha3*, proposed) for seed hardness on chromosome 16 (Fig. 7). This putative QTL associated with seed hardness lies in an approximately 46 cM region closely linked to markers Satt414 and Satt547 (Fig. 7, Table 3) and explained 32% of phenotypic variations with a LOD score of 2.8. Alignment of these markers was in agreement with public map (Song et al., 2004) indicating that

the genomic sequence of the population is in the same pattern as the original public mapping. These linked markers, Satt414 and Satt547, are 28 cM apart in the QTL region. Unfortunately markers located between these two markers were not polymorphic in this population, hampering the effort for fine mapping of this hardness QTL. Therefore, SNP markers will be needed in this region to examine allelic diversity and fine map the QTL for seed hardness. However, it is possible the new QTL could be distal in position to either Satt414 or Satt547. For example, marker Sat_224 is located at 75 cM (approx. 7 cM downstream from Satt547) and explained 5% of phenotypic variation with $P \leq 0.05$ value. The amount of variation explained by each marker on chromosome 16 varied and decreased with the downstream position. Therefore, SNP analysis should be done for the upstream positions to find out the exact location of this new QTL and whether there are more than one QTL in this region.

Previously reported seed hardness QTL, *Ha1* and *Ha2*, are located on chromosomes 19 and 1, respectively (Zhang et al., 2008). In the present study, Satt254 on chromosome 1 was consistent in three of four environments and located between Satt531 (at 40.9 cM) and Sat_110 (at 62.5 cM), both of which were reported to be linked to the hardness QTL *Ha2*. Therefore, this QTL region is confirmed in our population. Satt229 that is linked to *Ha1* (Zhang et al., 2008) was not polymorphic in our population and other markers on this chromosome were not associated with seed hardness in this study. Evidently, this QTL was not present in our population or under the environmental conditions in this study. Although eight chromosomal regions had association with seed hardness yet not all regions were stable across environments. For example, chromosomes 11, 13, and 18 had markers associated with seed hardness at two of four environments suggesting further research as potential regions containing other hardness QTL.

It is worth noting that markers (Satt414 and Satt547) linked to newly identified seed hardness QTL *Ha3* on chromosome 16 were also significantly associated with calcium content (Chapter 4). Moreover, Satt292 on chromosome 20 exhibited significance for both calcium content and hardness in present study (Chapter 4). Satt677 on chromosome 7 was also reported to be linked to calcium content QTL *Ca3* and *Ca4* (Zhang et al., 2009) supporting the observed correlation between calcium content and seed hardness. Finding common regions for calcium content and seed hardness suggests the existence of QTL with pleiotropic effect on calcium content and seed hardness endorsing the proposed idea of indirect selection using one trait/marker for another.

Comparison of marker allele effects on seed hardness in the 120 RIL showed differences ranging from 21 to 72 N between lines with hard-seeded parent PI 407818 B allele and lines with soft seeded parent KS4303sp allele, depending on the year and location (Table 4). The combined data across locations and years showed a 27 to 46 N difference between the soft and hard parent alleles. A total of seven chromosomal regions were associated with seed hardness; some regions contributed more difference in seed hardness than others. On average, each marker allele would theoretically contribute 36 N in hardness, while the difference between the two parents was 189 N and the range of the population was 310 N. Obviously, effects of some of the marker alleles were additive in contributing the seed hardness differences in the 120 RIL population. In addition, the possibility for epistatic effects of marker alleles cannot be ruled out. Future research is needed to investigate specific contribution of QTL and marker alleles to the phenotype.

Chromosome regions found to be associated with seed hardness contained markers associated with other important traits such as amino acid content and disease resistance. For example, Satt254 on chromosome 1, near QTL *Ha2*, linked to amino acid content QTL *trp*1-3,

*glu*1-2, *leu*1-1, *asp*1-2, and *ile*1-3 (Panthee et al., 2006). Moreover, Satt292 on chromosome 20 was reported to be linked to QTL for *trp* and *asp*. In addition, markers on chromosome 16 Satt547, linked to the new seed hardness QTL (*Ha3*), and Sat_224 were previously reported for soybean cyst nematode resistance. Evidently, the chromosome regions for seed hardness contain a cluster of genes/QTL for other important traits such as amino acid content and disease resistance. It would be an interest to study the relationship between seed hardness and amino acid content.

In conclusion, SSR markers and QTL confirmed and newly identified in the present study will improve efficiency in breeding for food-grade soybean lines with softer seeds and better texture. Confirmed QTL (*Ha2*) region on chromosome 1 (Zhang et al., 2008), a new QTL *Ha3* on chromosome 16, and markers Satt677 on chromosome 7, and Satt292 on chromosome 20 are the best candidates to be used for seed hardness selection. Exhibited additive allelic effect of markers associated with seed hardness in the present study need to be investigated for specific contribution of QTL to phenotype. These QTL/markers for seed hardness provide important information for food-grade soybean variety development with desired texture for natto production.

REFERENCES

Argel, P.J. and C.J. Paton. 1999. Overcoming legume hardseededness. In Forage seed production: Tropical and subtropical species (D.S. Loch and J.E. Ferguson, ed.) (Vol. 2, pp. 247-276). CAB International, Wallingford.

Calero, E., S.H. West, and K. Hinson. 1981. Water absorption of soybean seeds and associated casual factors. Crop Sci. 21:926-933.

Chen. P., M.R. Huhn, G.R. Buss, I. Gunduz, and J.H. Wilson. 2001. Interrelationship among seed quality traits of specialty soybeans for natto and tofu. ASA-CSSA-SSSA International annual meeting. Oct. 21-25, Charlotte, NC.

Cornelious, B., P. Chen, Y. Chen, D. Leon, J.G. Shannon, and D. Wang. 2005. Identification of QTLs underlying water-logging tolerance in soybean. Mol. Breed. 16:103–112.

Cregan, P.B., T. Jarvik, A.L. Bush, R.C. Shoemaker, K.G. Lark, A.L. Kahler, N. Kaya, T.T. VanToai, D.G. Lohnes, J. Chung, and J.E. Specht. 1999. An integrated genetic linkage map of the soybean genome. Crop Sci. 39: 1464-1490.

Hill, H.J., S.H. West, and K. Hinson. 1986. Effect of water stress during seedfill on impermeable expression in soybean. Crop Sci. 26:807-812.

Keim, P., B.W. Diers, and R.C. Shoemaker. 1990. Genetic analysis of soybean hard seededness with molecular markers. Theor. Appl. Genet. 79:465-469.

Kisha, T., C.H. Sneller, and B.W. Diers. 1997. Relationship between genetic distance among parents and genetic variance in populations of soybean. Crop Sci. 37:1317–1325.

Kosambi, D.D. 1944. The estimation of map distance from recombination values. Ann. Eugen. 12:172–175.

Kuo, W.H.J. 1989. Delayed permeability of soybean seeds: Characteristics and screening methodology. Seed Sci. and technology 27:131-142.

Ma, F., E. Cholewa, T. Mohamed, C.A. Peterson, and M. Gijzen. 2004. Cracks in the palisade cuticle of soybean seed coat correlate with their permeability to water. Ann. Bot. 94:213-228.

Mullin, W.J. and W. Xu. 2001. Study of soybean seed coat components and their relationship to water absorption. J. Agri. and Food Chem. 49 (11): 5331-5335.

Panthee, D.R., V.R. Pantalone, A.M. Saxton, D.R. West, and C.E. Sams. 2006. Genomic regions associated with amino acid composition in soybean. Mol. Breed. 17:79-89.

Qutob, D., F. Ma, C.A. Peterson, M.A. Bernards, and M. Gijzen. 2008. Structural and permeability properties of the soybean seed coat. Bot. 86: (3) 219-227.

Ragus, L.N. 1987. Role of water absorbing capacity in soybean germination and seedling vigour. Seed Science and Technology 15:285-296.

Sakamoto, S., J. Abe, A. Kanazawa, and Y. Shimamoto. 2004. Marker-assisted analysis for soybean hard seededness with isozyme and simple sequence repeat loci. Breeding Sci. 54(2):133-139.

SAS Institue. 2002. SAS/STAT User's guide, Version 9.6 ed. SAS Institute Inc. Cary, NC.

SAS Institute, 2006: JMP user?s guide, Version 6.0. SAS Institute Inc., Cary, NC.

Shao, S., C.J. Meyer, F. Ma, C.A. Peterson, and M.A. Bernards. 2007. The outermost cuticle of soybean seeds: Chemical composition and function during imbibitions. J. Exp. Bot. 58:1071-1082.

Song, J., G. An, and C. Kim. 2003. Color, texture, nutrient contents, and sensory values of vegetable soybeans [*Glycine max* (L.) Merrill] as affected by balancing. Food Chem. 83:69-74.

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

Taira, H. 1990. Quality of soybeans for processed foods in Japan. Japan Agric. Res. Quarterly 24:224-230.

Tinius, C. 1991. Variable expression of seedcoat permeability. Soybean Genetics Newsletter 18:256-259.

Van Ooijen, J.W., and R.E. Voorrips. 2001. JoinMap version 3.0, software for the calculation of genetic linkage maps. Plants Research Int., Wageningen, the Netherlands.

Voorrips, R.E. 2002. Mapchart: Software for the graphical presentation of linkage maps and QTLs. J. Hered. 93:77–78.

Wang, D., J. Shi, S.R. Carlson, P.B. Cregan, R.W. Ward, and B.W. Diers. 2003. A low-cost, high-throughput polyacrylamide gel electrophoresis system for genotyping with microsatellite DNA markers. Crop Sci. 43:1828–1832.

Watanabe, S., T. Tajuddin, N. Yamanaka, M. Hayashi, K. Hayashi, and K. Harada. 2004. Analysis of QTLs for reproductive development and seed quality traits in soybean using recombinant inbred lines. Breed. Sci. 54:399-407.

Werker, E., I. Marbach and A. Mayer. 1979. Relation between the anatomy of the testa, water permeability, and the presence of phenolics in the genus *Pisum*. Ann. Bot. 43**:** 765-771.

Yalkich, R.W., W.P. Wergin, and E.L. Vigil. 1986. Special secretory cells in the soybean seed coat. Protoplasma 134:78-87.

Zhang, B. (2006). Genetic analysis of food-grade soybeans and identification of quantitative trait loci (QTL) for seed hardness and calcium content in soybean.

Zhang, B., P. Chen, A. Shi, A. Hou, T. Ishibashi, and D. Wang. 2009. Putative quantitative trait loci associated with calcium content in soybean seed. J. Hered. 100(2):263-269.

Zhang. B., P. Chen, C.Y. Chen, D. Wang, A. Shi, A. Hou, and T. Ishibashi. 2008. Quantitative trait loci mapping of seed hardness in soybean. Crop Sci 48:1341–134.

Table 1. Analysis of variance for seed hardness of 120 RIL from the cross KS4303sp x PI 407818 B grown at Fayetteville, AR and Keiser, AR in 2008 and 2009.

Source	Degrees of freedom	Sum of squares	Mean square	F-value	P-value	R2
Model	488	3873959.8	7938.4	1.83	<0.0001	0.70
Year	1	93315.6	93315.6	21.54	<0.0001	
Location	1	4971.2	4971.2	1.15	0.28	
Year*Location	1	521798.7	521798.7	120.45	<0.0001	
Genotype	119	1375311.1	11273	2.6	<0.0001	
Year*Genotype	119	910750.6	7589.6	1.75	<0.0001	
Location*Genotype	119	413981	3393.3	0.78	0.90	
Year*Location*Genotype	119	483703.4	4134.2	0.95	0.60	
Error	392	1698170.9	4332			

Table 2. Summary of SSR markers used in screening parents and polymorphic markers used in screening the RIL population derived from the cross KS4303sp x PI 407818 B.

Chromosome	No. SSR markers screened	Coverage (cM)	No. polymorphic markers	Coverage (cM)
1	25	102.96	9	68.27
2	27	137.06	9	80.25
3	21	76.21	4	65.39
4	18	127.77	0	0
5	25	101.57	7	83.39
6	25	125.08	5	110.24
7	24	112.79	9	82.79
8	35	136.13	8	126.21
9	23	115.22	2	45.71
10	20	124.36	4	62.48
11	20	91.62	3	67.84
12	20	100.86	2	17.99
13	30	140.43	8	117.27
14	23	107.56	5	40.93
15	22	68.01	2	23.65
16	19	78.59	4	30.12
17	22	104.21	3	48.38
18	40	108.7	9	95.47
19	25	115.07	5	91.35
20	22	82.28	7	59.94
Total	**486**	**2156**	**105**	**1318**
Average	**24.30**	**107.82**	**5.25**	**12.50**

Table 3. Single marker analysis of variance for seed hardness of 120 RIL population from the cross KS4303sp x PI 407818 B grown in Fayetteville, AR (FAY) and Keiser, AR (KEI) in 2008 and 2009.

		2008		2009		Combined†	
Chromosome	**Position (cM)**	**FAY**	**KEI**	**FAY**	**KEI**		**R^2**
1	17.52					**Satt184‡**	
1	40.87					**Satt531‡**	
1	56.43	Satt254*		Satt254*	Satt254*	Satt254*	0.05
1	62.52					**Sat_010‡**	
7	66.99			Satt175***	Satt175		
7	75.57	Satt677		Satt677	Satt677*	Satt677**	0.08
7	76.41			Sat_288	Sat_288	Sat_288*	0.05
11	73.77			Satt597***	Satt597***	Satt597**	0.08
13	3.63			Satt586	Satt586		
13	135.12			Satt656**	Satt656**	Satt656**	0.07
16	37.04	Satt414**	Satt414**	Satt414		Satt414****	0.15
16	38.19	Satt406**	Satt406**			Satt406***	0.12
16	67.79	Satt547		Satt547***	Satt547	Satt547***	0.11
16	75.13			Sat_224*	Sat_224*	Sat_224*	0.05
18	52.94			Satt594*		Satt594**	0.07
18	57.32		Satt564*	Satt564**			
19	93.89					**Satt229‡**	
20	75			Sat_170**		Sat_170*	0.05
20	82.78	Satt292		Satt292**	Satt292	Satt292**	0.08

† refers to the pooled data from two locations in two years.

‡ refers to the markers previously reported for seed hardness (Zhang et al., 2008).

*, **, ***, and **** represent significant association with seed hardness at P ≤0.05, 0.01, 0.001, and 0.0001, respectively.

Table 4. Mean effect of SSR marker alleles on seed hardness (in newtons) in 120 RIL population from the cross KS4303sp x PI 407818 B in 2008 and 2009.

SSR/Chromosome		2008			2009			Combined*			
		PI 407818 B allele	KS4303sp allele	Allelic difference	PI 407818 B allele	KS4303sp allele	Allelic difference	PI 407818 B allele	KS4303sp allele	Allelic difference	R
Satt254	1	365	336	**29**	363	323	**40**	364	330	**35**	0.05
Satt175	7	399	364	**35**	367	330	**37**	383	347	**36**	
Sat_288	7	384	347	**37**	368	340	**28**	376	344	**33**	0.05
Satt677	7	389	344	**45**	367	332	**35**	378	338	**40**	0.08
Satt597	11	379	350	**29**	377	328	**49**	378	339	**39**	
Satt586	13	378	354	**24**	368	339	**29**	373	347	**27**	
Satt656	13	368	347	**21**	367	333	**34**	368	340	**28**	0.07
Satt406	16	416	354	**62**	366	342	**24**	391	348	**43**	0.12
Sat_224	16	391	350	**41**	369	341	**28**	380	346	**35**	0.05
Satt547	16	387	348	**39**	368	325	**43**	378	337	**41**	0.11
Satt414	16	413	344	**69**	362	340	**22**	388	342	**46**	0.15
Satt594	18	369	327	**42**	366	322	**44**	368	325	**43**	0.07
Satt564	18	382	356	**26**	367	336	**31**	375	346	**29**	
Sat_170	20	382	347	**35**	374	339	**35**	378	343	**35**	0.05
Satt292	20	378	345	**33**	369	332	**37**	374	339	**35**	0.08

*refers to the pooled data from two years and two locations.

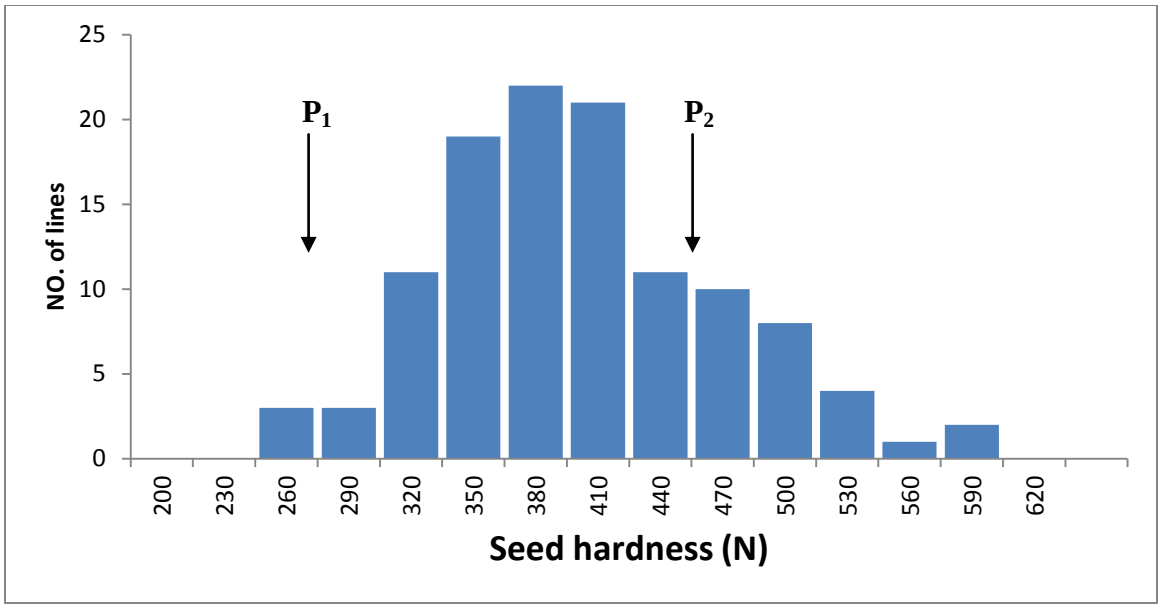


Figure 1. Frequency distribution of seed hardness of 120 RIL from the cross KS4303sp (P1) x PI 407818 B (P2), grown in Fayetteville, AR in 2008.

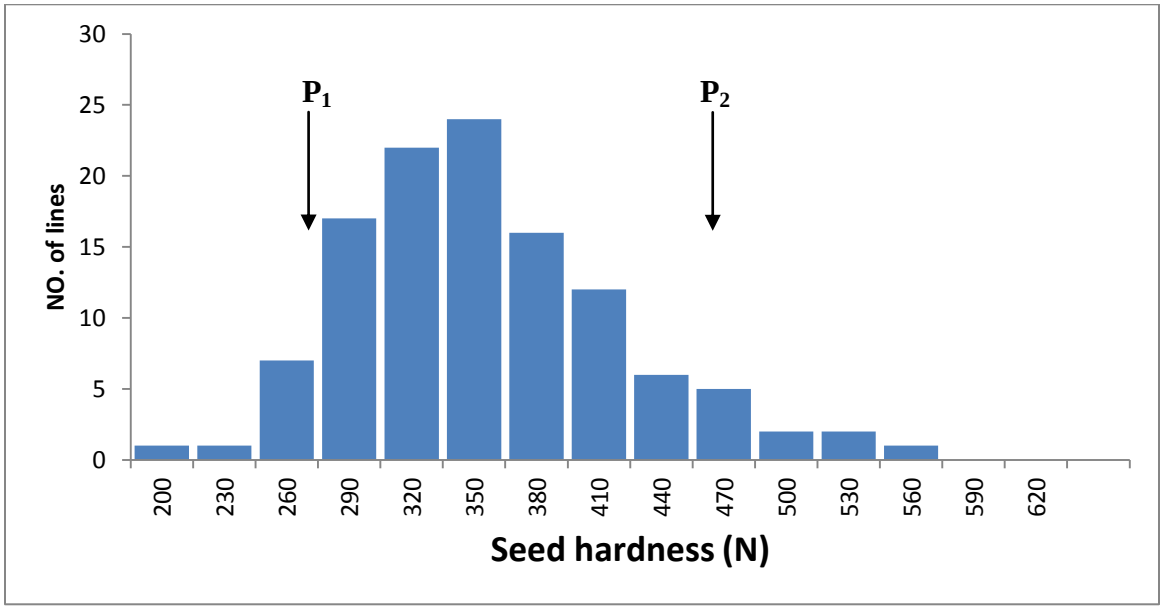


Figure 2. Frequency distribution of seed hardness of 120 RIL from the cross KS4303sp (P1) x PI 407818 B (P2), grown in Keiser, AR in 2008.

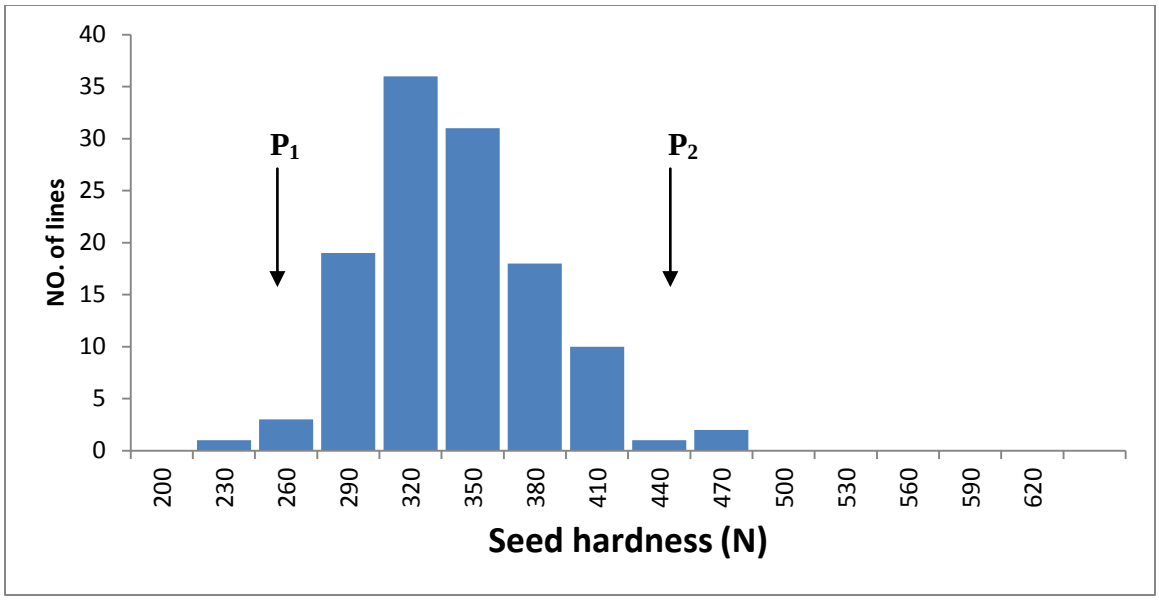


Figure 3. Frequency distribution of seed hardness of 120 RIL from the cross KS4303sp (P1) x PI 407818 B (P2), grown in Fayetteville, AR in 2009.

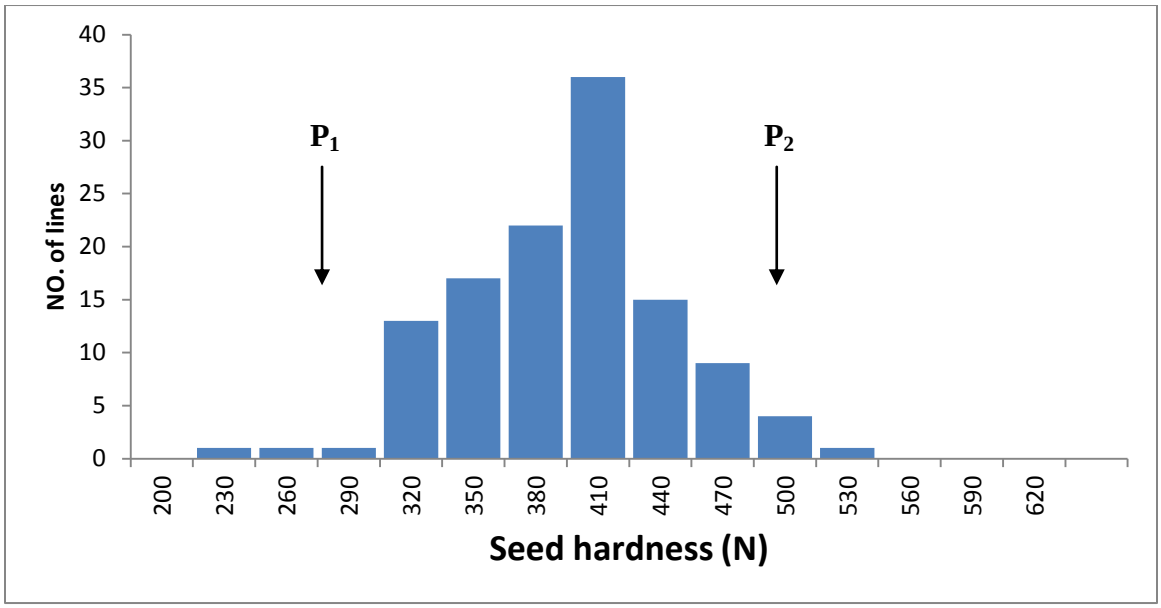


Figure 4. Frequency distribution of seed hardness of 120 RIL from the cross KS4303sp (P1) x PI 407818 B (P2), grown in Keiser, AR in 2009.

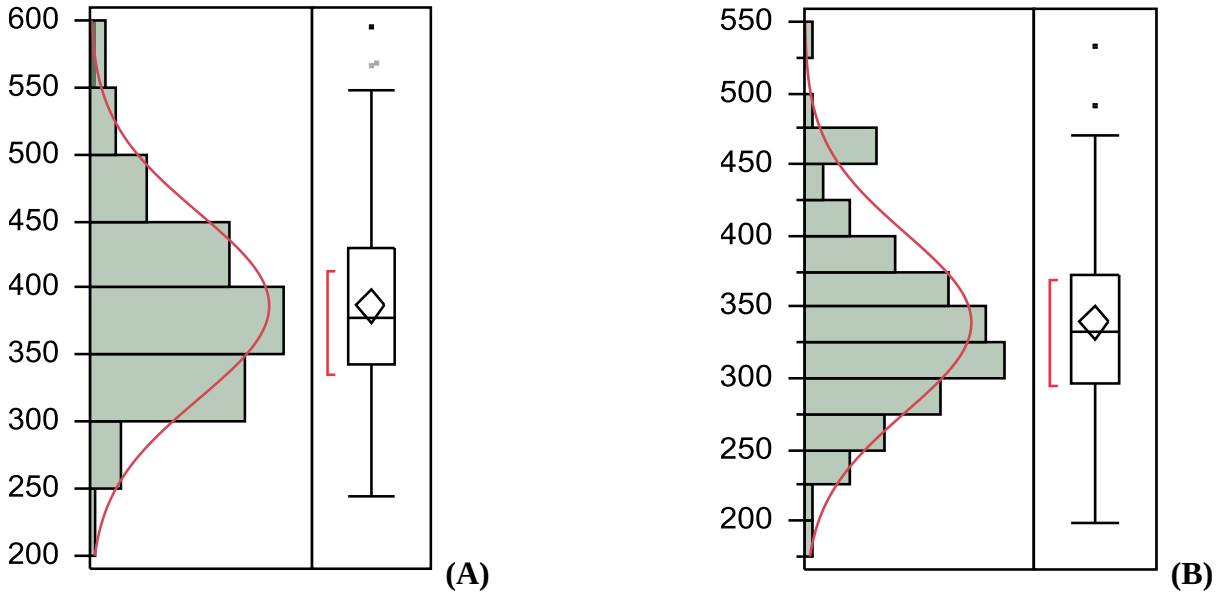


Figure 5. Normal distribution of seed hardness of 120 RIL from the cross KS4303sp x PI 407818 B, grown in Fayetteville, AR (A) and in Keiser, AR (B) in 2008.

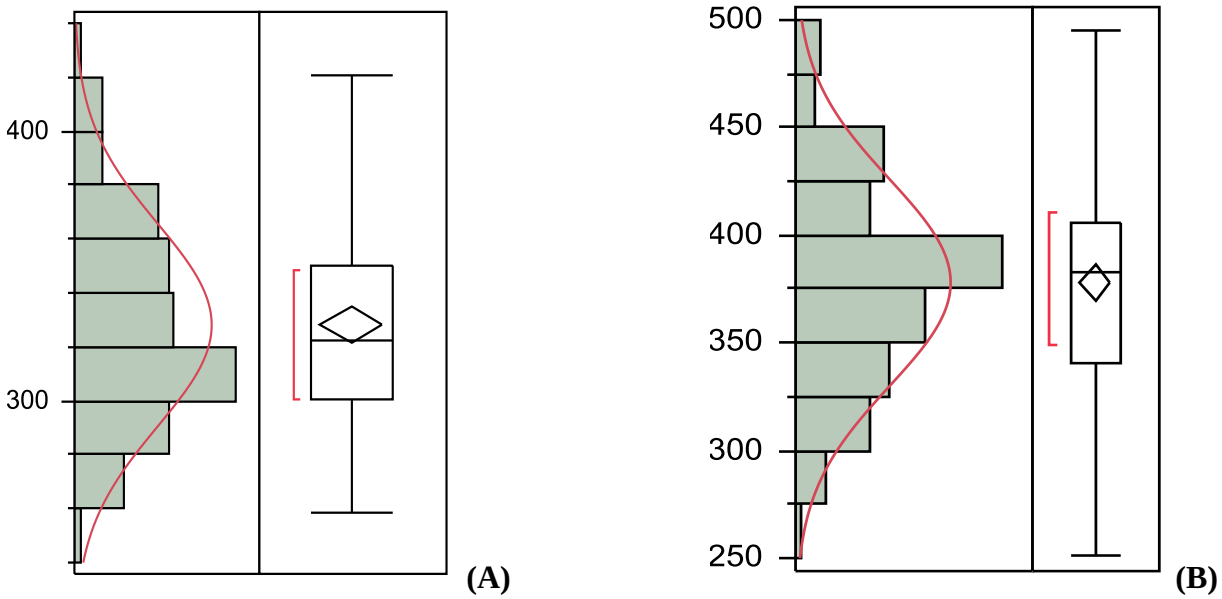


Figure 6. Normal distribution of seed hardness of 120 RIL from the cross KS4303sp x PI 407818 B, grown in Fayetteville, AR (A) and in Keiser, AR (B) in 2009.

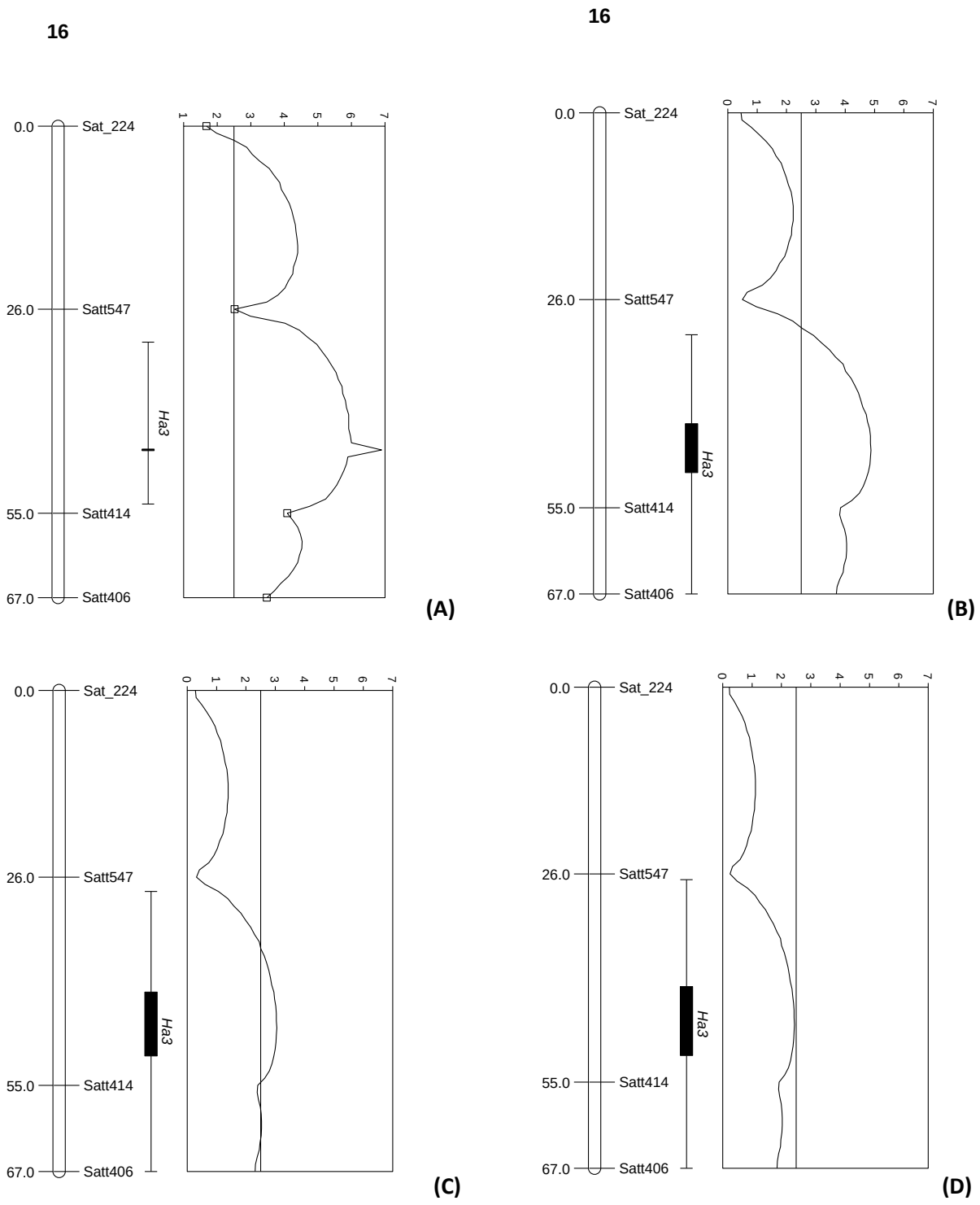


Figure 7. Interval mapping for soybean seed hardness on chromosome 16 in 120 RIL derived from the cross KS4303sp x PI 407818 B at Fayetteville (A) and Keiser (B) in 2008 and Fayetteville (C) and Keiser (D) in 2009.

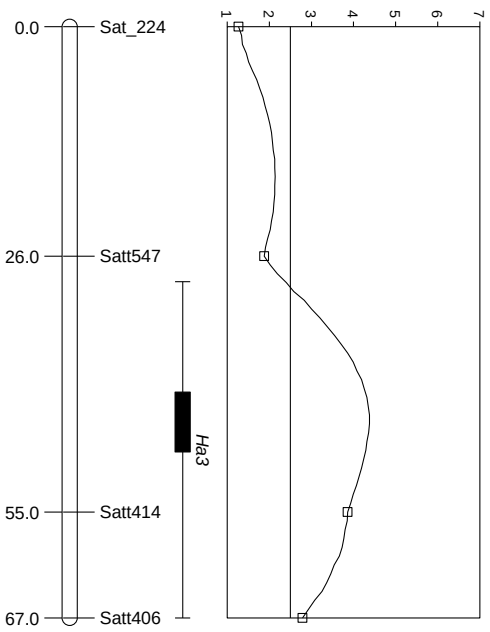


Figure 8. Interval mapping for soybean seed hardness on chromosome 16 in 120 RIL derived from the cross KS4303sp x PI 407818 B using combined data across locations (Fayetteville and Keiser, AR) and the years (2008 and 2009).

Chapter IV

Confirmation of SSR Markers and QTL for Seed Calcium Content and Hardness of Soybean and Correlation of These Traits

ABSTRACT

Development of food-grade soybean cultivars with desired traits is essential due to the demands for soyfood production. Seed calcium content and hardness are main determining characteristics of soybean texture affecting soyfood quality. Molecular markers closely linked to these traits will accelerate the process of breeding specialty soybeans for the soyfood market through the use of marker assisted selection (MAS). Confirmation of linked markers and validation of QTL before applying to MAS are important steps. The objectives of this study were to confirm previously reported QTL for seed calcium content and hardness and investigate the relationship between these two traits. Evaluation of seven recombinant inbred line (RIL) populations with different genetic backgrounds grown at two locations in Arkansas for two years showed inconsistent correlation between these traits. However, depending on the environments positive correlation was found in most of the RIL populations with six being significant. The combined data showed consistently positive correlation between calcium content and seed hardness. Previously reported QTL for seed calcium content and/or hardness were evaluated in 6 different sub-populations and stable markers over years and locations in different genetic backgrounds were identified as a potential tool for MAS. Markers Satt267 and Sat_345 on chromosome 1, Sat_288 on chromosome 7, Sat_228, Satt341, and Sat_392 on chromosome 8, Satt547 on chromosome 16, and Satt002 on chromosome 17 will be recommended for calcium content selection; while Satt267 on chromosome 1, Satt680 on chromosome 7, Satt341 on chromosome 8, and Sct_010 on chromosome 19 will be used for seed hardness selection. Findings of this research will facilitate MAS for seed calcium content and hardness in soybean breeding programs for food-grade soybeans.

INTRODUCTION

Traditional oriental soyfood makes almost 90% of the food quality soybean market (Taira H., 1990). Food grade soybeans require certain traits that affect soyfood quality and consumer acceptance. The quality of tofu and natto is affected by the characteristics of the seed utilized in their production. Food-grade varieties should have specific characteristics such as yellow colored seed, clear hilum, small or large soft seeds with high sucrose and low calcium content. Natto production requires small seeds yet small seeds tend to have higher calcium content which is not desirable for natto. Seeds with higher calcium content have harder texture which is another undesirable character for natto processing. Breeders face challenges when high yield and desired seed traits need to be combined in one cultivar in order to meet the need of the expanding soyfood market. Combination of conventional breeding and molecular approaches may help efficiently assemble combinations of desired genes in a new variety.

Development of molecular markers in plant breeding was introduced in late 1900th. Since then molecular markers have been used to improve the efficiency of crop breeding. Recent advances in molecular breeding using marker technology improved the understanding of genetics of important traits of soybean such as protein, oil, and sugar. The main purpose of understanding the genetics of the traits is to be able to assemble more desirable combinations of genes in new varieties. Many traits that affect quality of soybean seed are quantitative in nature. Using molecular markers closely linked to the target QTL is used in marker assisted selection (MAS). The role of MAS is important in crop breeding. It has many advantages over conventional breeding such as early selection of the trait in a seedling stage and flexibility of selection environments. It is also important for the complex traits that require time consuming and labor intensive analysis because it eases phenotypic screening. DNA fingerprinting before planting or

harvesting reduces the number of plants to grow or harvest so that plants with target traits can be selected. It also enables gene pyramiding where more than one trait can be assembled or where multiple alleles control the trait. Although MAS has many advantages, it is essential to perform reliable research for linked molecular markers before applying. Markers associated with QTL from preliminary mapping studies cannot be used directly in MAS. Confirmation and validation of QTL before applying to MAS are important steps (Langridge et al., 2001). Confirming and validating QTL and linked markers in different genetic backgrounds are essential for their implications in different populations. In addition, validation of QTL/linked markers in different environments is also crucial for the use of MAS in different breeding programs.

Calcium content in the seed coat affects its water absorption and texture. Chen et al., (2001) indicated that calcium content in the seed was positively correlated with seed hardness after cooking and both traits varied due to environmental effects such as temperature and soil type. Since calcium has a significant role in cell membrane, it could affect water absorption by making cell wall thicker or having calcium oxalate cell layer on soybean seed wall. Having molecular markers for the calcium content would ease the process of calcium content analysis which is time consuming and labor intensive. As the genetics of this trait applies to specialty soybeans for soyfood production, molecular markers and QTL, once identified and confirmed, can be used for improvement of the micronutrient concentration to obtain high calcium varieties for nutrient considerations or varieties with low calcium where texture of the product is important.

High temperature, low air relative humidity, and water stress cause harder seeds (Argel and Paton, 1999). Depending on weather conditions, amount of hard seed vary with year and geographic locations (Calero et al., 1981). Genetic factors, geographic location, and weather

conditions during plant growth determine the amount of hard seeds; however, permeability of seeds is primarily determined by genotype (Tinius, 1991). Soybean seed hardness and its role in texture is an important considerations in soyfood production. Information on genetic control of calcium content and seed hardness will be useful in selecting soybean lines for specialty soybeans for soyfood market.

Correlation between soybean seed calcium content and seed hardness was reported previously and higher calcium content can be the cause of harder seeds. Zhang et al., (2008) indicated that small seeds tend to have higher calcium content and consequently, cause harder seeds. Research also showed that larger seeds tend to have lower calcium content and be softer when they are cooked (Zhang et al., 2008, Chen at al., 2001). Taira (1990) also showed that seed size is negatively correlated with seed water absorption and cooked seed hardness. This association of small seed size and seed hardness has been observed in other research with soybean and other plant species (Yaklich et al., 1986, Ragus 1987). Soybean seed hardness and calcium content and their role in texture are important in soyfood production and difficult to measure. Therefore, understanding the genetics and correlation of these two traits will facilitate the process of selection for soybean lines with desirable calcium content and seed hardness, thereby, improving the efficiency of soybean breeding process for specialty soybeans for soyfood market.

Objectives of this study were to investigate correlation between seed calcium content and hardness, confirm previously reported QTL for these traits in different genetic backgrounds and environments, and identify reliable markers for MAS.

MATERIALS AND METHODS

Plant materials and field experiment

A total of 164 advanced breeding lines derived from six crosses with different genetic backgrounds were chosen for QTL confirmation purposes: 31 lines derived from Hutcheson x Camp, 30 lines derived from SS-516 x Camp, 20 lines derived from SS-516 x V96-4486, 20 lines derived from Hutcheson x V96-4486, 20 lines derived from MFS-591 x Camp, and 43 lines derived from MFS-591 x V96-4486. These lines were selected based on their performance on agronomic traits such as seed size, yield, lodging, and height for natto soybean breeding purposes. All six crosses involved small seeded parents with contrasting differences in seed calcium content and hardness. These lines were selected for agronomic traits only and assumed to represent a random population on the basis of seed calcium content and hardness. Therefore, they were chosen to examine the correlation between these traits and confirm previously reported QTL. The selected lines were grown in Fayetteville and Kibler, AR in 2007 and Fayetteville and Keiser, AR in 2008 with 2 replications with randomized complete block design (RCBD). Parental lines were included as checks in the tests.

One hundred twenty recombinant inbred lines (RIL) derived from KS4303sp x PI 408078 B used for mapping for calcium content and seed hardness were also used to investigate correlation between these two traits. These 120 RIL were grown in Fayetteville and Keiser, AR in 2008 and 2009 with RCBD. Parental lines were included as checks in the test. All small-seeded breeding lines and mapping RIL were grown in a single row plot with 3 m in length and 0.95 m row spacing. Irrigation was conducted according to University of Arkansas irrigation schedule. Once plants in a row reached 95% maturity, they were bulk harvested either with a

combine or a plant thresher. Field plots were managed using standard cultural practices adopted for full season soybean production in Arkansas.

DNA extraction

Leaf tissues were collected from the uppermost partly developed trifoliate leaf of plants in each line. The bulked tissue samples were ground with a mortar and pestle in liquid nitrogen and stored in a -20^{o}C freezer. Total genomic DNA was isolated using CTAB (hexadercyltrimethyl ammonium bromide) method (Kisha et al., 1997). Briefly, 0.50 mg powder obtained from ground leaves were transferred to a 2.0 ml microtube and 750 µl of CTAB extraction buffer was added followed by incubation at 65^{o}C for 60 min. inverting every 10 min. Tubes were cooled and 500 µl chloroform:isoamyl alcohol (24:1) was added and inverted several times followed by centrifuging samples at 13,500 rpm for 15 min. The aqueous layer was transferred to a 2.0 ml microtube and centrifuged for 5 min. which lead to DNA precipitation. DNA was washed with 95% ethanol solution. Supernatant was poured out and DNA pellets were allowed to air dry at room temperature overnight. 250 µl of 0.1 x TE was added over dried DNA and left in temperature for a day to dissolve DNA. Each sample was checked for its concentration using Bio-Tek PowerWave XS Microplate Spectrophotometer.

SSR polymorphism analysis

Polymerase chain reaction (PCR) and simple sequence repeat (SSR) primer genotyping were performed in a 96-well or 384-well BIO-RAD iCycler thermocycler (Bio-Rad, Hercules, CA). PCR reaction was performed using 4 µl of 20ng/µl template DNA, 1.0 µM of 0.5 µM forward and reverse primers, 0.2 µl of Go Taq Flexi polymerase, 0.9 µl of 2.5 mM dNTP (nucleotides) mixture, 1.8 µl of 2.5 mM $MgCl_2$, and 4.3 deionized H_2O. PCR reaction were at 94^{o}C for 4 min. followed by 33 cycles at 94^{o}C for 25 s for denaturation, 47^{o}C for 25 s for

annealing, 68°C for 25 s for extension, and 72°C for 5 min. for final extension after the last cycle. Amplified PCR products were separated on a 6% non-denaturing polyacrylamide gel on a Megagel electrophoresis system (C.B.S. Scientific) and stained with ethidium bromide (Wang et al., 2003). Previously reported SSR markers and QTLs were used to screen these advanced breeding lines (Table 5). Gels were visualized under UV fluorescence and digital images were taken for scoring individuals. Primers for the SSR markers used in this study, their sequences, and their integration in soybean molecular linkage map are publicly available online at http://129.186.26.94/ssr.html (Cregan et al., 1999, Song et al., 2004).

Calcium analysis

Calcium content was determined by HNO_3 method (Campbell and Plank, 1991). Ten grams of seeds from each plot were ground by Krups KM75 coffee grinder until fine powder was obtained. 0.25 grams of powder were digested by 2.5 ml HNO_3. Samples were gradually heated to 60°C for 45 min. and slowly increased to 120°C for an hour after adding 1 ml H_2O_2. Cooled samples were mixed with deionized water and filtered through #41 quantitative paper followed by calcium content analysis using Spectro Ciros with simultaneous inductively coupled plasma (Spectro Analaytical Instruments, Inc., Mahwah, NJ).

Hardness test

Thirty grams of unbroken uniform seeds from each line was weighed and soaked in heat resistant plastic boxes containing 150 mL water at ambient temperature for 16 hours. Seeds were recovered from soaking water with a sieve and blot dried on paper towels. Stone seeds that did not absorb water during soaking process were removed from soaked seeds. Soaked seed samples were pressure cooked in heat resistant plastic boxes at 121.1°C and 1.2 kg cm^{-2} for 20 min. Hardness tests of 30 g cooked seeds from each entry were conducted in two replications using a

TMS Texture System (TMS-2000, Food Technology Corp., Sterling, VA) equipped with a 16-blade shear cell. The maximum force to compress cooked seeds in newtons (N) was determined as seed hardness (Song et al., 2003).

QTL analysis

Association between molecular data and phenotypic data used to confirm SSR markers and QTL were analyzed by single-factor analysis of variance with the PROC GLM procedures of SAS software v9.2 (SAS Institute, 2002). WinQTL cartographer (Basten et al., 1999) was used for single marker analysis (SMA) to identify QTL associated with calcium content. For SMA, $p<0.05$ was used as a threshold for significant markers (Cornelious et al., 2005

Statistical data analysis

Phenotypic data collected in 2007 and 2008 were analyzed separately and combined with the PROC MEANS procedure in SAS software v9.2 (SAS Institute, 2002). The PROC GLM procedure was used to determine variance components for traits with location and replication as random blocking factors. Correlation was estimated using CORR procedure in SAS for pair-wise combination of traits from the RIL means at each location and in each year prior to the combined data analysis.

RESULTS AND DISCUSSION

Correlation between seed calcium content and hardness

Relationship between seed calcium content and hardness was investigated using two genetic populations. One population includes 120 RIL derived from KS4303sp x PI 407818 B with parents having contrasting seed calcium content and hardness (Table 1, Fig. 1 and 2). The other population incudes 164 breeding lines derived from six different crosses, in all cases both parents exhibited contrasting differences in both seed calcium content and hardness (Tables 3

and 4; Fig. 3 and 4). These 164 breeding lines were selected for agronomic traits including yield, plant height, seed size, lodging, and maturity. However, for seed calcium content and hardness, these lines represented a random population and were used for confirming QTL/molecular markers and examining correlation between the two traits.

For the RIL population from KS4303sp x PI 407818 B, the correlation analysis showed a positive, but weak association between calcium content and seed hardness at both locations in 2008. However, in 2009 a significantly positive correlation was observed between seed calcium content and hardness for both locations (r = 0.21 - 0.23). The overall analysis showed a positive correlation (r = 0.28) between seed calcium content and hardness for the combined data from both years and locations (Table 1). For the 164 breeding line population, the correlation coefficient was variable among crosses and across environments. In most cases, the correlation between seed calcium content and hardness was positive with some being statistically significant. Three crosses (SS-516 x Camp, MFS 591 x Camp, and MFS 591 x V96-4486) exhibited consistent and positive correlation in two or three environments and combined data for each cross across environments also showed a significantly positive correlation (r = 0.41 - 0.49). The other three crosses showed inconsistent correlation across environments, but the combined data across environments indicated a positive (r = 0.23 - 0.39), although not statistically significant, correlation between seed calcium content and hardness (Table 2). This instability of the correlation between seed calcium content and hardness due to environmental conditions was also reported in previous research (Chen et al., 2001). However, in the present study all six sub-populations showed a positive correlation between seed calcium and hardness with three populations being statistically significant (r = 0.23 – 0.49). Combined analysis of all six sub-populations across environments indicated a strong positive correlation between seed calcium

content and hardness ($r = 0.38$ and $P<0.0001$). This trait association was also observed in other studies (Chen etal., 1993, 2001). Evidently, in a practical breeding program, dual selections for both calcium content and seed hardness can be achieved by measuring only one of the two traits whichever is easier to evaluate. The indirect selection for both traits can also be applied successfully to the MAS when reliable markers for one trait are identified and confirmed. This would improve the efficiency of a breeding program in which food-grade quality attributes are an objective.

Frequency distribution of seed calcium content and hardness

For the ease of presentation, data on seed calcium content and hardness were combined across locations due to the fact that location and genotype x location effects were not significant in the ANOVA (data not shown). Both seed calcium content and hardness were normally distributed in both years for the KS4303sp x PI 407818 B population (Fig. 1 and 2). The pooled data from both years showed that seed calcium content ranged from 0.20 to 0.30% and hardness ranged from 260 to 520 N. Low calcium/soft seeded parent KS4303sp had seed calcium content of 0.20% and seed hardness of 260 N, while the high calcium/hard-seeded parent PI 407818 B had a calcium content of 0.31% and hardness of 470 N.

All six sub-populations derived from high calcium/hard seed x low calcium/soft seed crosses segregated for both traits as expected with ranges and means being slightly different in magnitude (Tables 3 and 4). Since the population size was small and unequal in numbers, all six sub-populations were combined as a single population for distribution analysis (Fig. 3 and 4). The combined data showed a normal distribution for both calcium content and seed hardness (Fig. 3 and 4). Calcium content from the pooled data ranged from 0.27 to 0.42% and seed hardness from 350 to 720 N. Low calcium/soft seeded parents had an average of 0.26% calcium

and 419 N for seed hardness, while high calcium/hard seeded parents had an average of 0.36% calcium and 535 N for seed hardness. It is worth noting that the calcium content was higher and seed harder in 2007 than 2008 due to different climate conditions. However, lines ranked similarly in both years. The normal distribution of all seven populations above indicate that both calcium content and seed hardness are typical quantitative traits conferred by QTL and influenced by environmental conditions and that they are representative and appropriate for confirming QTL/markers for both traits.

Confirmation of SSR markers and QTL for seed calcium content and hardness

Based on the integrated soybean linkage map (Cregan et al., 1999), 59 polymorphic out of 237 total SSR markers closely located to the potential QTL regions for calcium content and seed hardness, identified in the present study (Chapters 2 and 3) and previous research (Zhang et al, 2008, 2009), were screened in the six sub-populations derived from high calcium/hard seed x low calcium/soft seed crosses (Table 5). Some of the previously reported markers were monomorphic in parental lines and therefore not used in the screen of breeding populations. Results showed that markers on eight out of 14 chromosomes were associated with seed calcium content across six sub-populations (Table 6). Among these validated markers and chromosome regions in the present populations, the ones on chromosomes 1, 8, 16, and 8 were confirmed in three or four sub-populations. In addition, QTL regions and markers on chromosomes 1, 7, 8, 16, and 17 were consistent across different environments. Markers Satt267 and Sat_345 on chromosome 1, Sat_288 on chromosome 7, Sat_228, Satt341, and Sat_392 on chromosome 8, Satt547 on chromosome 16, and Satt002 on chromosome 17 appeared to be most consistent across sub-populations and environments, hence the best choices of markers for MAS for

calcium content. It is possible that these stable markers (across different genetic backgrounds and environments) will help improve efficiency of selection for lines with proper calcium content.

It is worth noting that Sat_288 on chromosome 7, Sat_228, Satt341, and Sat_392 on chromosome 8 are closely located to the two previously reported QTL regions (*Ca1*, *Ca3* and *Ca4*) for calcium content (Zhang et al., 2009). It is also important to point out that Sat_290 on chromosome 18 is near the new QTL identified in this research (*Ca5*, see chapter 2) for calcium content and was present in three out of six sub-populations. Interestingly, markers Satt267 and Sat_345 on chromosome 1 are also close to the QTL (*Ha2*) for hardness (Zhang et al. 2008), while Satt547 on chromosome 16 is the same marker for the new hardness QTL (*Ha3*) identified in the present research (Chapter 3), supporting the observed correlation between calcium content and seed hardness and the proposed concept for indirect selection using one trait/maker for another. Other chromosome regions/markers, although not associated with identified QTL (Zhang et al., 2009, chapter 2) but associated with calcium content in the six sub-populations and the 120 RIL from KS4303sp x PI 407818 B, could imply potential new QTL for calcium content and deserve further research in the future.

Eight of 14 chromosome regions had markers associated with seed hardness across six sub-populations (Table 7). Markers on chromosome 1, 7, and 8 were confirmed in five of six sub-populations and markers on chromosome 16, 17, 19, and 20 were confirmed in three or two of six sub-populations for seed hardness. Moreover, QTL regions and markers on chromosome 1, 7, 8, 17, 19, and 20 were consistent across environments. Considering the frequency of markers confirmed for seed hardness in different sub-populations and environments, Satt267 on chromosome 1, Satt680 on chromosome 7, Satt341 on chromosome 8, and Sct_010 on chromosome 19 are the best candidates for MAS for seed hardness. In addition, Satt267 on

chromosome 1 and Sct_010 on chromosome 19 are located close to two previously reported QTL, *Ha2* and *Ha1* (Zhang et al., 2008), respectively, confirming the importance of these regions for seed hardness. It is important to point out that Satt414 on chromosome 16 confirmed in two of the six sub-populations was shown to be the same marker for the new hardness QTL *Ha3* identified in the present research (Chapter 3). Two of the best markers for seed hardness, Satt680 on chromosome 7 and Satt341 on chromosome 8, were also associated with seed calcium content in the present study, supporting the observed correlation between these two traits. Other markers on chromosomes 16, 17, 18, and 20 associated with seed hardness, yet not consistent in different environments and genetic backgrounds, may contain new QTL and need additional research in the future.

Findings of present study show that some regions of the chromosome are responsible for one trait while other regions may impact both calcium content and seed hardness. It is not surprising that some QTL have pleiotropic effect on both calcium and seed hardness, given that fact of close correlation of both traits. However, it is possible that unique QTL exist, either for calcium content or seed hardness, which could be genetic background specific and/or environment dependent.

Chromosome regions found to be associated with calcium content and/or seed hardness contain markers reported to be linked to some other traits such as seed amino acid content and resistance to diseases including cyst nematode, corn earworm, stem rot, and common cutworm (Panthee et al., 2006, Komatsu et al., 2005, Arahana et al., 2001, Li et al., 2010, and Mahalingam and Skorupska, 1995). For example, markers on chromosome 1 which are associated with both seed calcium content and hardness are located in the similar region where seed tryptophan, leucine, isoleucine, and glutamic and aspartic acid content QTL reside (Panthee et al., 2006).

Markers on chromosome 7 for both calcium and seed hardness are in the similar region where amino acid and pest resistance QTL for common cutworm (*Spodoptera Litura*) are located (Komatsu et al., 2005). On chromosome 8, region at 77 to 116 cM where several markers were identified for seed calcium content and hardness also contain previously reported QTL for seed amino acid (proline, alanine, and serine) content (Panthee et al., 2006) and QTL for pest resistance such as soybean cyst nematode (*Heterodera glycines*) (Mahalingam and Skorupska, 1995), Sclerotinia stem rot (*Sclerotinia sclerotiorum*) (Arahana et al., 2001), and *Phyophtora sojae* infection (Li et al., 2010). On chromosome 16, markers located between 37 to 67 cM region for calcium content and seed hardness are close to the region associated with resistance to soybean cyst nematode (*Heterodera glycines*) (Glover et al., 2004) and *Phyophtora sojae* infection (Weng et al., 2007). On chromosome 18, the new QTL for calcium content and seed hardness in present study is located close to the markers associated with seed amino acid content (tryptophan, leucine, and cysteine) (Panthee et al., 2006), soybean cyst nematode (*Heterodera glycines*) (Guo et al., 2005, Yue et al., 2001), *Sclerotinia* stem rot (*Sclerotinia sclerotiorum*) (Arahana et al., 2001). On chromosome 19, markers for calcium content and seed hardness are closely located to QTL identified previously for seed amino acid content (serine, glycine, proline, alanine, threonine, and valine) (Panthee et al., 2006) and resistance to *Sclerotinia* stem rot (*Sclerotinia sclerotiorum*) (Arahana et al., 2001). It is likely that these chromosome regions contain clusters of genes for different traits such as seed composition (such as calcium and amino acid) and disease resistance/stress tolerance. It was suggested that cell membrane and cell wall strength have an important role in pest control. Calcium is an important element of the cell wall and most distributed in the soybean seed coat. Therefore, the calcium metabolism may have a role to play in plant defense system. It is possible that hard seed with high calcium content and

less permeability may be less susceptible to pathogen attack. Such hypothesis needs further investigation in future research.

In conclusion, confirmed QTL/markers reported previously and new QTL/markers identified in the present study offer a potential tool for MAS for calcium content and seed hardness. Markers Satt267 and Sat_345 on chromosome 1, Sat_288 on chromosome 7, Sat_228, Satt341, and Sat_392 on chromosome 8; Satt547 on chromosome 16, and Satt002 on chromosome 17 are the best candidates to be used for calcium content selection; while Satt267 on chromosome 1, Satt680 on chromosome 7, Satt341 on chromosome 8, and Sct_010 on chromosome 19 for seed hardness selection. As these QTL/markers appear to be stable across genetic backgrounds and environments, it is hopeful that they can be adopted in any breeding programs for food-grade soybean variety development and that the application of these markers in a MAS scheme will accelerate selection of progeny lines with desired seed attributes.

REFERNCES

Arahana, V.S., G.L. Graef, J.E. Specht, J.R. Steadman, and K.M. Eskridge. 2001. Identification of QTLs for Resistance to *Sclerotinia sclerotiorum* in Soybean. Crop Sci. 41(1):180-188.

Argel, P.J. and C.J. Paton. 1999. Overcoming legume hardseededness. In Forage seed production: Tropical and subtropical species (D.S. Loch and J.E. Ferguson, ed.) (Vol. 2, pp. 247-276). CAB International, Wallingford.

Basten, C.J., B.S. Weir, and Z.B. Zeng. 1999. QTL cartographyer, version 1.13. Dep. of Statistics, North Carolina State Univ., Raleigh.

Calero, E., S.H. West, and K. Hinson. 1981. Water absorption of soybean seeds and associated casual factors. Crop Sci. 21:926-933.

Campbell, C.R. and C.O. Plank. 1991. Sample preparation in plant analysis reference procedures for the southern region of the United States. Southern Coop. Serv. Bull. 368:1–11.

Chen. P., M.R. Huhn, G.R. Buss, I. Gunduz, and J.H. Wilson. 2001. Interrelationship among seed quality traits of specialty soybeans for natto and tofu. ASA-CSSA-SSSA International annual meeting. Oct. 21-25, Charlotte, NC.

Cornelious, B., P. Chen, Y. Chen, D. Leon, J.G. Shannon, and D. Wang. 2005. Identification of QTLs underlying water-logging tolerance in soybean. Mol. Breed. 16:103–112.

Cregan, P.B., T. Jarvik, A.L. Bush, R.C. Shoemaker, K.G. Lark, A.L. Kahler, N. Kaya, T.T. VanToai, D.G. Lohnes, J. Chung, and J.E. Specht. 1999. An Integrated Genetic Linkage Map of the Soybean Genome. Crop Sci. 39: 1464-1490.

Glover, K.D., D. Wang, P.R. Arelli, S.R. Calson, S.R. Cianzio, and B.W. Diers. 2004. Near isogenic lines confirm a soybean cyst nematode resistance gene from PI 88788 on linkage group J. Crop Sci. 44(3):936-941.

Guo, B., D.A. Sleeper, P.R. Arelli, J.G. Shannon, and H.T. Nguyen. 2005. Identification of QTLs associated with resistance to soybean cyst nematode races 2, 3 and 5 in soybean PI 90763. Theor. Appl. Genet. 103(8):1167-1173.

Kisha, T., C.H. Sneller, and B.W. Diers. 1997. Relationship between genetic distance among parents and genetic variance in populations of soybean. Crop Sci. 37:1317–1325.

Komatsu, K., S. Okuda, M. Takahashi, R. Matsunaga, and Y. Nakazawa. 2005. QTL mapping of antibiosis resistance to common cutworm (*Spodoptera litura fabricius*) in soybean. Crop Sci. 45(5):2044-2048.

Kosambi, D.D. 1944. The estimation of map distance from recombination values. Ann. Eugen. 12:172–175.

Langridge, P., E. Lagudah, T. Holton, R. Appels, P. Sharp, and K. Chalmers. 2001. Trends in

genetic and genome analyses in wheat: a review. Augt. J. Agric. 52:1043-1077.

Li, X., Y. Han, W. Teng, S. Zhang, K. Yu, V. Poysa, T. Anderson, J. Ding, and W. Li. 2010. Pyramided QTL underlying tolerance to *Phytophthora* root rot in mega-environments from soybean cultivars Conrad and Hefeng 25. Theor. Appl. Genet. 121(4): 651-658.

Mahalingam, R. and H.T. Skorupska. 1995. DNA markers for resistance to *Heterodera glycines* I. race 3 in soybean cultivar Peking. Breeding Sci. 45(4):435-443.

Panthee, D., V. Pantalone, A. Saxton, D. West, and C. Sams. 2006. Genomic regions associated with amino acid composition in soybean. Mol. Breed. 17(1):79-89.

Ragus, L.N. 1987. Role of water absorbing capacity in soybean germination and seedling vigour. Seed Science and Technology 15:285-296.

Reinprecht, Y., V. Poysa, K. Yu, I. Rajcan, G. Ablett, and K. Pauls. 2006. Seed and agronomic QTL in low linolenic acid, lipoxygenase-free soybean (*Glycine max* (L.) Merrill) germplasm. Genome. 49(12):1510-1527.

SAS Institue. 2002. SAS/STAT User's guide, Version 9.6 ed. SAS Institute Inc. Cary, NC.

SAS Institute, 2006: JMP user's guide, Version 6.0. SAS Institute Inc., Cary, NC.

Song, J., G. An, and C. Kim. 2003. Color, texture, nutrient contents, and sensory values of vegetable soybeans [*Glycine max* (L.) Merrill] as affected by balancing. Food Chem. 83:69-74.

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

Taira, H. 1990. Quality of soybeans for processed foods in Japan. Japan Agric. Res. Quarterly 24:224-230.

Tinius, C. 1991. Variable expression of seedcoat permeability. Soybean Genetics Newsletter 18:256-259.

Van Ooijen, J.W. and R.E. Voorrips. 2001. JoinMap version 3.0, software for the calculation of genetic linkage maps. Plants Research Int., Wageningen, the Netherlands.

Voorrips, R.E. 2002. Mapchart: Software for the graphical presentation of linkage maps and QTLs. J. Hered 93:77–78.

Wang, D., J. Shi, S.R. Carlson, P.B. Cregan, R.W. Ward, and B.W. Diers. 2003. A low-cost, high-throughput polyacrylamide gel electrophoresis system for genotyping with microsatellite DNA markers. Crop Sci. 43:1828–1832.

Weng, C., K. Yu, T. Anderson, and V. Poysa. 2007. A quantitative trait locus influencing

tolerance in *Phytophthora* root rot in the soybean cultivar 'Conrad'. Euph. 158(1-2):81-86.

Yalkich, R.W., W.P. Wergin, and E.L. Vigil. 1986. Special secretory cells in the soybean seed coat. Protoplasma 134:78-87.

Yue, P., P. Arelli, and D.A. Sleper. 2001. Molecular characterization of resistance to *Heterodera glycines* in soybean PI 438489B. Theor. Appl. Genet. 102(6-7):921-928

Zhang, B., P. Chen, A. Shi, A. Hou, T. Ishibashi, and D. Wang. 2009. Putative quantitative trait loci associated with calcium content in soybean seed. J. Hered. 100(2):263-269.

Zhang. B., P. Chen, C.Y. Chen, D. Wang, A. Shi, A. Hou, and T. Ishibashi. 2008. Quantitative trait loci mapping of seed hardness in soybean. Crop Sci 48:1341–134.

Table 1. Pair-wise analysis of Pearson's correlation coefficient for average calcium content and seed hardness in a 120 RIL population derived from the cross KS4303sp x PI 407818 B evaluated at Fayetteville, AR (FAY) and Keiser, AR (KEI) in 2008 and 2009.

FAY 08	KEI 08	FAY 09	KEI 09	Combined†
0.15	0.03	0.23**	0.21*	0.28**

† refers to the pooled data from four environments.

* and ** refer to significance of correlation coefficient at $P \leq 0.05$ and 0.01, respectively.

Table 2. Pair-wise analysis of Pearson's correlation coefficient for average seed calcium content and hardness in 164 breeding lines derived from six crosses evaluated at Fayetteville, AR (FAY) in 2007 and Fayetteville, AR and Keiser, AR (KEI) in 2008.

Cross	No. of lines	FAY 07	FAY 08	KEI 08	Combined†
Hutcheson x Camp	31	-0.08	0.39*	-0.29	0.27
SS516 x Camp	30	0.29	0.59***	-0.006	0.49**
SS-516 x V96-4486	20	-0.001	0.36	0.07	0.23
Hutcheson x V96-4486	20	0.32	-0.05	0.018	0.39
MFS-591 x Camp	20	0.38	0.54**	-0.01	0.47*
MFS-591 x V96-4486	43	0.35**	0.2	0.31*	0.41**
Overall	164	0.20*	0.33****	0.09	0.38****

† refers to the pooled data from three environments (1 location in 2007 and 2 locations in 2008).

*, **,***, and **** refer to the significance at $P \leq 0.05$, 0.01, 0.001, and 0.0001, respectively.

Table 3. Average seed calcium content of parental lines and 164 breeding lines derived from six crosses evaluated in two locations (Fayetteville, AR and Keiser, AR) and over years (2007 and 2008).

		Low calcium parent		High calcium parent		Population	
Cross	**No. of lines**	**Range**	**Mean**	**Range**	**Mean**	**Range**	**Mean**
Hutcheson x Camp	31	0.24-0.27	0.26	0.33-0.36	0.35	0.22-0.43	0.32
SS516 x Camp	30	0.22-0.29	0.26	0.33-0.36	0.35	0.21-0.42	0.31
SS-516 x V96-4486	20	0.22-0.29	0.26	0.30-0.40	0.37	0.23-0.46	0.32
Hutcheson x V96-4486	20	0.24-0.27	0.26	0.30-0.40	0.37	0.25-0.50	0.33
MFS-591 x Camp	20	0.21-0.28	0.25	0.33-0.36	0.35	0.23-0.41	0.32
MFS-591 x V96-4486	43	0.21-0.28	0.25	0.30-0.40	0.37	0.23-0.49	0.32
Average			0.26		0.36	0.23-0.45	0.32

Table 4. Average seed hardness of parental lines and 164 breeding lines derived from six crosses evaluated in two locations (Fayetteville, AR) in 2007 and (Fayetteville, AR and Keiser, AR) in 2008.

		Soft-seeded parent		Hard-seeded parent		Population	
Cross	**No. of lines**	**Range**	**Mean**	**Range**	**Mean**	**Range**	**Mean**
Hutcheson x Camp	31	379-547	444	437-670	551	211-806	496
SS516 x Camp	30	367-439	402	437-670	551	285-830	477
SS-516 x V96-4486	20	367-439	402	444-614	519	255-750	478
Hutcheson x V96-4486	20	379-547	444	444-614	519	288-764	481
MFS-591 x Camp	20	391-477	412	437-670	551	294-868	507
MFS-591 x V96-4486	43	391-477	412	444-614	519	303-731	488
Average			419		535	272-791	488

Table 5. Summary of polymorphic SSR markers used in screening 164 breeding lines derived from six crosses (Hutcheson x Camp, SS516 x Camp, SS-516 x V96-4486, Hutcheson x V96-4486, MFS-591 x Camp, and MFS-591 x V96-4486) for association analysis.

Chromosome	Position (cM)	SSR marker	Calcium	Hardness
1	17.52	Satt184		+
1	36.23	Sat_353		+
1	40.90	Satt531		+
1	48.14	Satt342		+
1	56.43	Satt254		+
1	57.34	Satt267		+
1	62.50	Sat_110		+
1	65.63	Sat_345		+
1	77.49	Satt077		+
1	108.89	Satt147		+
2	25.60	Satt095	+	
2	72.89	Satt141	+	
2	84.04	Satt041	+	
5	42.80	Sat_356	+	
7	1.02	Sat_391	+	
7	33.47	Satt567	+	
7	66.99	Satt175	+	+
7	75.57	Satt677	+	+
7	76.41	Sat_288	+	+
7	77.19	Satt680	+	+
7	86.86	Satt123		+
7	95.45	Satt551		+
7	118.14	Satt153		+
7	112.79	Satt346		+
8	77.70	Satt341	+	+
8	84.09	Sat_199	+	
8	90.84	Satt377	+	
8	106.29	Sat_392	+	+
8	116.41	Sat_382	+	+
8	116.64	Sat_377	+	
8	116.73	Satt470	+	
8	154.11	Satt228	+	
9	32.96	Satt055	+	
11	73.77	Satt597		+

12	53.35	Satt541	+	
13	3.63	Satt586		+
13	3.95	Satt030		+
13	16.08	Satt252	+	
13	135.12	Satt656	+	
16	37.04	Satt414	+	+
16	38.19	Satt406	+	+
16	67.79	Satt547	+	+
17	47.73	Satt002	+	
17	80.19	Satt461	+	
18	18.25	Satt217	+	
18	27.48	Sat_315	+	
18	29.03	Sat_290	+	
18	43.78	Satt115	+	
18	45.49	Sat_358	+	
18	52.94	Satt594	+	
18	57.32	Satt564	+	
18	87.20	Satt564	+	
19	34.54	Satt313		+
19	59.52	Sct_010		+
20	36.59	Sat_174	+	
20	46.22	Satt354	+	
20	75.00	Sat_170	+	
20	77.48	Satt330	+	
20	82.78	Satt292	+	

Table 6. Single marker analysis of variance for seed calcium content of 164 breeding lines derived from six crosses evaluated in Fayetteville, AR (FAY) and Kibler, AR (KIB) in 2007 and Fayetteville, AR and Keiser, AR (KEI) in 2008.

			2007		2008		Combined†
Cross	Chromosome	Position (cM)	FAY	KIB	FAY	KEI	
Hutcheson x Camp	1	57.34		Satt267**	Satt267**		
	1	65.63					Sat_345*
	16	67.79			Satt547*		
SS-516 x Camp	1	57.34		Satt267**	Satt267**		Satt267*
	1	65.63	Sat_345**	Sat_345**	Sat_345**		Sat_345*
	7	1.02	Sat_391**	Satt391**			
	7	76.41	Sat_288*	Sat_288**	Sat_288***		Sat_288**
	7	77.19	Satt680**				
	8	77.7	Satt341**	Satt341***	Satt341***		Satt341***
	8	106.29	Sat_392**	Sat_392**	Sat_392**		Sat_392***
	8	116.41		Sat_382**			Sat_382*
	16	37.14	Satt414**		Satt414**		Satt414*
	16	67.79	Satt547*	Satt547*			
	18	29.03				Sat_290*	
	19	59.52			Sct_010*		
SS-516 x V96-4486	1	57.34		Satt267**			
	16	67.79	Satt547*				
	20	75	Sat_170*				
	20	77.48	Satt330**				Satt330*
Hutcheson x V96-4486	1	65.63	Sat_345*		Sat_345*		Sat_345**
	8	77.7		Satt341*			

	16	67.79		Satt547**			
MFS-591 x Camp	18	29.03				Sat_290**	
	19	59.52		Sct_010**	Sct_010*		Sct_010*
MFS-591 x V96-4486	8	116.41			Sat_382*		Sat_382*
	8	154.11	Satt228*	Satt228*			Satt228*
	17	47.73	Satt002*	Satt002**	Satt002*		Satt002**
	18	29.03		Sat_290*			
	20	75	Sat_170**	Sat_170*			Sat_170*

† refers to the pooled data from two locations in two years.

*, **, and *** represent significant association with calcium content at P ≤0.05, 0.01, and 0.001, respectively.

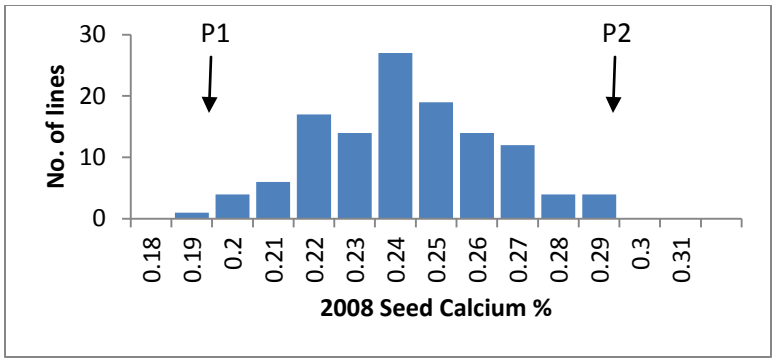
Table 7. Single marker analysis of variance for seed hardness of 164 breeding lines derived from six crosses evaluated in Fayetteville, AR (FAY) in 2007 and Fayetteville, AR and Keiser, AR (KEI) in 2008.

			2007	2008		Combined†
Cross	Chromosome	Position (cM)	FAY	FAY	KEI	
Hutcheson x Camp	1	57.34				Satt267**
	7	77.19			Satt680*	Satt680*
	7	95.45	Satt551*			
	8	106.29	Sat_392*			
SS-516 x Camp	1	57.34		Satt267**	Satt267***	Satt267**
	7	76.41		Sat_288***	Sat_288***	Sat_288**
	7	77.19		Satt680**	Satt680**	
	8	77.7		Satt341**	Satt341**	
	8	106.29		sat_392*		
	16	37.04		Satt414*		
	17	47.73		Satt002**	Satt002**	Satt002*
	19	59.52		Sct_010**	Sct_010**	Sct_010*
	20	75		Sat_170*	Sat_170*	
	20	77.48		Satt330**		
SS-516 x V96-4486	1	57.34			Satt267*	
	7	112.79			Satt346*	Satt346*
	8	77.7			Satt341*	
	8	106.29			Sat_392*	
	17	47.73	Satt002*			
	19	34.54				Satt313*
	19	59.52	Sct_010**			

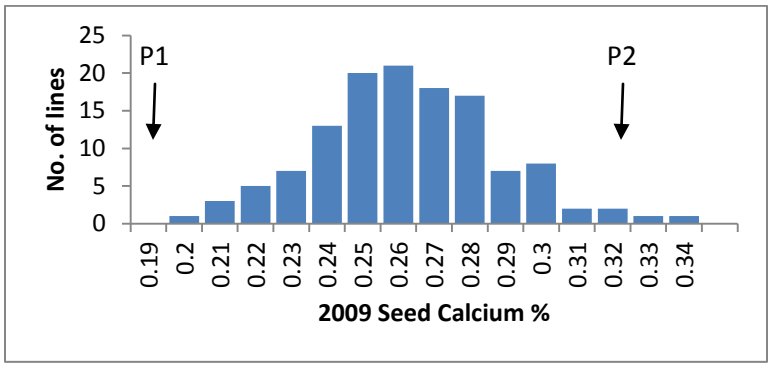
Hutcheson x V96-4486	1	57.34		Satt267**		
	1	65.63			Sat_345*	
	7	95.45		Satt551*		
	8	77.7	Satt341*			
	16	37.04	Satt414**			Satt414*
MFS-591 x Camp	1	65.63		Sat_345*		
	7	77.19		Satt680*		
	7	112.79		Satt346**		
	8	116.41		Sat_382**		
MFS-591 x V96-4486	18	43.78		Satt115*		
	19	59.52	Sct_010**			Sct_010*
	20	75	Sat_170*			Sat_170*

† refers to the pooled data from three environments.

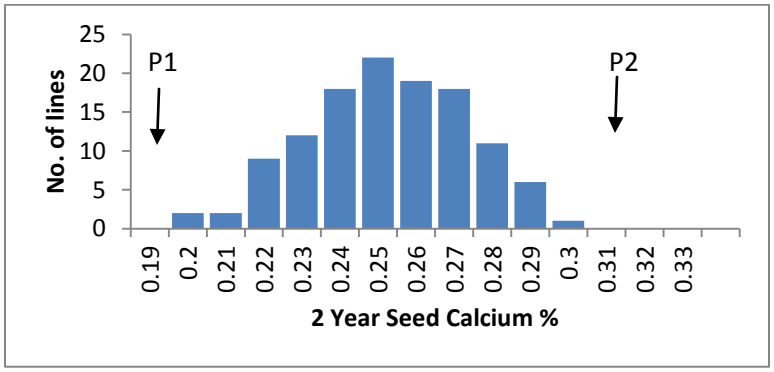
*, **, and *** represent significant association with calcium content at P≤ 0.05, 0.01, and 0.001, respectively.



(A)

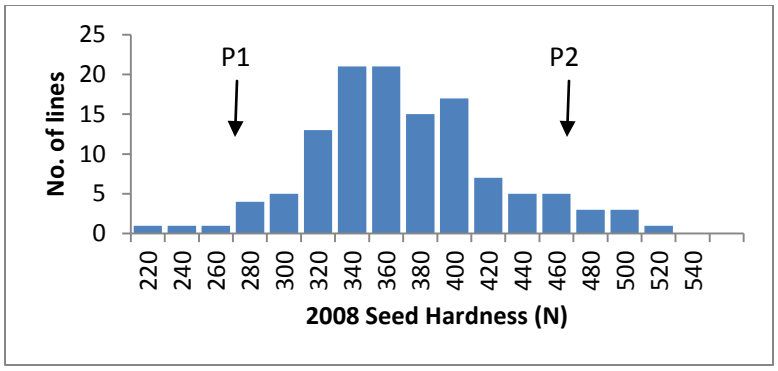


(B)

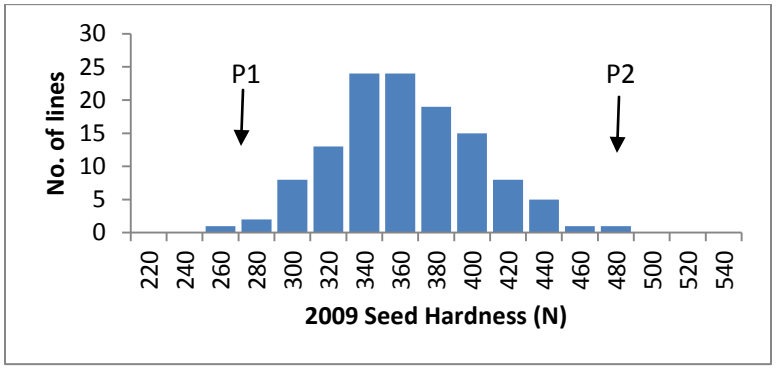


(C)

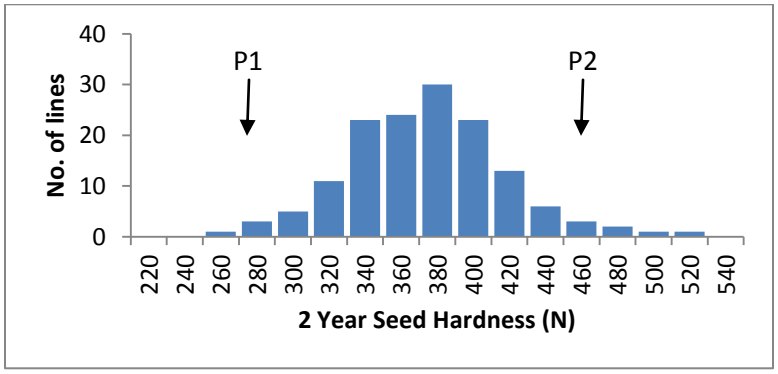
Figure 1. Frequency distribution of seed calcium content in 120 RIL derived from KS4303sp x PI 407818 B in 2008 (A), 2009 (B), and combined (C).



(A)

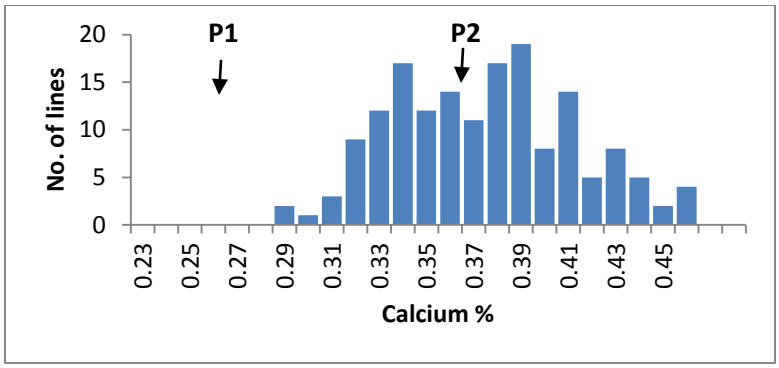


(B)

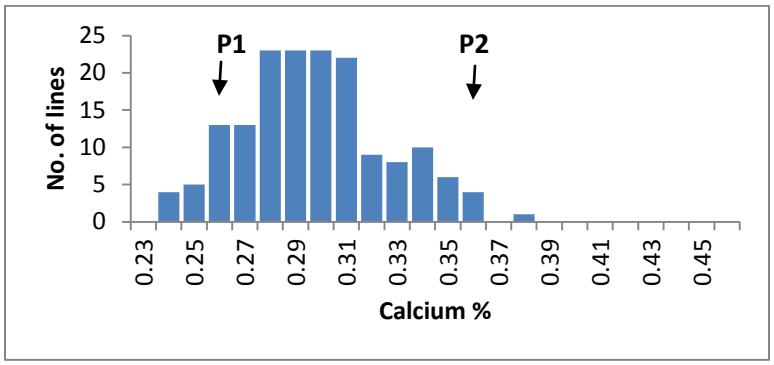


(C)

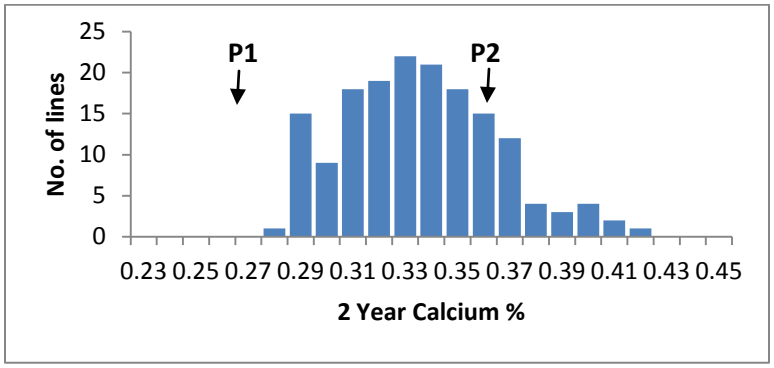
Figure 2. Frequency distribution of seed hardness in 120 RIL derived from KS4303sp x PI 407818 B in 2008 (A), 2009 (B), and combined (C).



(A)

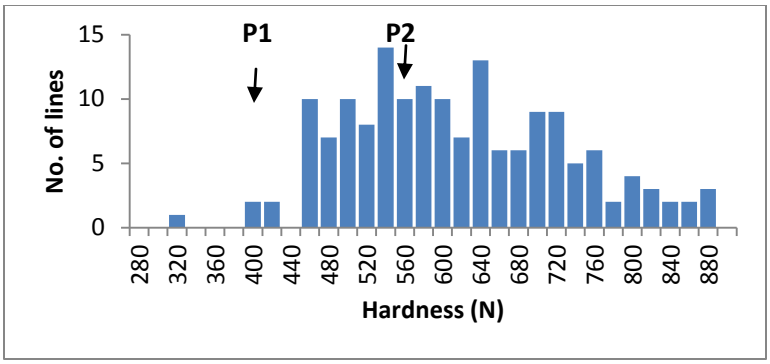


(B)

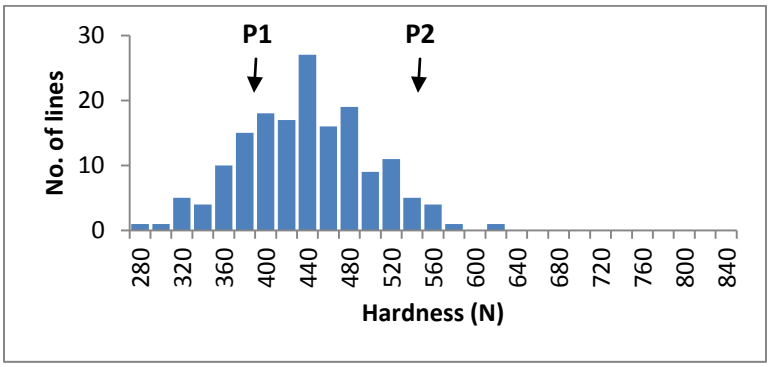


(C)

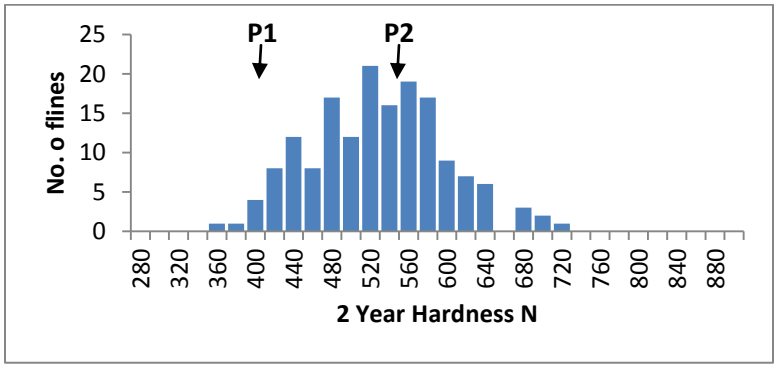
Figure 3. Frequency distribution of seed calcium content in 164 lines derived from six crosses evaluated in 2007 (A), 2008 (B), and combined (C).



(A)



(B)



(C)

Figure 4. Frequency distribution of seed hardness in 164 lines derived from six crosses evaluated in 2007 (A), 2008 (B), and combined (C).

Chapter V

CONCLUSION

Development of food grade soybean cultivars with desired traits is essential due to the demand in domestic and international soyfood market for products with high nutritional value. Seed calcium content and cooked hardness are determining factors in texture, hence an important attribute for food grade soybean varieties. Natto qualities are based not only on taste but also on texture, therefore understanding the genetics of traits affecting seed texture is important. Mapping populations consisting of 120 recombinant inbred lines RIL derived from KS4303sp (low calcium) x PI 407818 B (high calcium) and 124 RIL derived from PI 408052 C (low calcium/soft seeded) x PI 407818 B (high calcium/hard seeded) were used to identify SSR markers and QTL associated with seed calcium content and hardness. One hundred sixty four advanced breeding lines derived from six different low calcium soft-seeded x high calcium hard-seeded crosses were used to confirm previously reported and newly identified QTL for seed calcium content and hardness. Population derived from KS4303sp x PI 407818 B and 164 breeding lines from six crosses with differences in both traits was used to identify correlation between seed calcium content and hardness.

A new QTL (*Ca5*) for seed calcium content was identified on chromosome 18 to be linked to markers Sat_290 and Satt115. In addition, regions with a cluster of markers associated with seed calcium content were found on chromosomes 1 (Satt184, Sat_353, Satt342, Satt254, Satt077, Sat_408, and Satt147), chromosome 2 (Satt095, Satt141, Satt600, and Satt941), chromosome 8 (Sat_199, Satt377, Sat_377, Satt470, and Sat_040), and chromosome 18 (Satt217, Sat_315, Sat_290, Satt115, Sat_358, Satt594, and Satt564). Although previously reported QTL (*Ca1-Ca4*) were not identified in these new mapping populations, markers linked to those QTL such as Satt677 (*Ca3* and *Ca4*) on chromosome 7, Sat_377 (*Ca1*) on chromosome 8, and

Sat_174 (*Ca2*) on chromosome 20 were consistently associated with seed calcium content in our mapping populations.

A new QTL (*Ha3*) for seed hardness was identified on chromosome 16 and linked to markers Satt547 and Satt414. It contributed to 32 % of variation in seed hardness. In addition, stable markers on chromosome 1 (Satt254), chromosome 7 (Satt677), and chromosome 20 (Satt292) were consistently associated with seed hardness across environments in the population derived from KS43030sp x PI 407818 B. Previously reported seed hardness QTL *Ha2* on chromosome 1 was confirmed in the mapping population in the present study.

The most stable seed calcium content and hardness markers and QTL across different genetic backgrounds and environments were identified using 164 advanced breeding populations from six crosses. Satt267 and Sat_345 on chromosome 1, Sat_288 on chromosome 7, Sat_228, Satt341, and Sat_392 on chromosome 8, Satt547 on chromosome 16, and Satt002 on chromosome 17 were found to be the most stable markers across different genetic backgrounds and environments for seed calcium content. Markers Sa267 on chromosome 1, Satt680 on chromosome 7, Satt341 on chromosome 8, Sct_010 on chromosome 19 were most stable markers for seed hardness. Interestingly, markers Satt680 on chromosome 7, Satt341 on chromosome 8, and Satt414 on chromosome 16 were associated with both traits exhibiting a pleiotropic effect on both seed calcium content and hardness.

Correlation between seed calcium content and hardness was investigated using populations with differences in seed calcium content and hardness. A positive correlation was observed between seed calcium content and hardness. Combined data analysis across environments showed a positive correlation in all populations and most of them were statistically significant.

Findings of the present study will facilitate MAS for seed calcium content and hardness in soybean breeding programs for food-grade soybean. A positive correlation between seed calcium content and hardness combined with markers with pleiotropic effect on both traits suggests a possibility of the indirect selection for efficient or both traits.