University of Arkansas, Fayetteville

ScholarWorks@UARK

Graduate Theses and Dissertations

5-2018

An Evaluation of Breeding, Agronomic, and Processing Methodologies of Vegetable Soybean (Edamame) to Increase Domestic Production in the United States Market

David Octor Moseley
University of Arkansas, Fayetteville

Follow this and additional works at: https://scholarworks.uark.edu/etd

 Part of the Agronomy and Crop Sciences Commons, Food Processing Commons, and the Plant Breeding and Genetics Commons

Citation

Moseley, D. O. (2018). An Evaluation of Breeding, Agronomic, and Processing Methodologies of Vegetable Soybean (Edamame) to Increase Domestic Production in the United States Market. *Graduate Theses and Dissertations* Retrieved from https://scholarworks.uark.edu/etd/2628

This Dissertation is brought to you for free and open access by ScholarWorks@UARK. It has been accepted for inclusion in Graduate Theses and Dissertations by an authorized administrator of ScholarWorks@UARK. For more information, please contact scholar@uark.edu, uarepos@uark.edu.

An Evaluation of Breeding, Agronomic, and Processing Methodologies of Vegetable Soybean (Edamame) to Increase Domestic Production in the United States Market

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy in Crop, Soil, and Environmental Sciences

by

David Octor Moseley
Texas A&M University
Bachelor of Science in Agronomy, 2003
University of Arkansas
Master of Science in Agricultural, Food, and Life Sciences, 2010

May 2018
University of Arkansas

This dissertation is approved for recommendation to the Graduate Council.

_____
Pengyin Chen, Ph.D.
Dissertation Director

_____
Jeremy Ross, Ph.D.
Committee Member

_____
Kris Brye, Ph.D.
Committee Member

_____
Terry Kirkpatrick, Ph.D.
Committee Member

_____
Ruben Morawicki, Ph.D.
Committee Member

Abstract

Edamame is a food-grade soybean (*Glycine max* (L.) Merrill) that is either harvested immature (R6 stage) or after plant maturity (R8 stage). At R6, the premium product will have crescent-shaped large green pods and gray pubescence. At R8, the seed will either have stayed green or will have turned yellow, black, or brown. Edamame is a healthy snack with a sweet flavor and firm texture. The edamame market is growing in the United States, creating a need for more adapted varieties. The genetic diversity is low among adapted large-seeded breeding lines. Finding diverse accessions will help develop larger and more adapted varieties. Harvesting edamame at the R6 stage is challenging, as the harvest window can be <5 days. Research is needed to help define edamame breeding, production, and processing strategies. The objectives of this dissertation were to: i) discover quantitative trait loci (QTL) controlling seed weight and size traits in edamame-type germplasm; and compare the available diversity to large-seeded breeding lines from the University of Arkansas, ii) estimate the harvest window at the R6 stage; and evaluate the effects of planting date and variety on pod weight and color, and iii) improve shelf-stable edamame products by evaluating pasteurization methods of high moisture edamame. A total of 343 accessions and 31 breeding lines were used to discover QTLs and compare diversity of seed weight and size traits. Three varieties were planted in 12 environments to observe the optimum harvest date and harvest window at the R6 stage. A commercial edamame variety and three breeding lines with green, black, and brown seed were pasteurized in an acidic brine. Genetically, there were two main groups among the 343 accessions, and the accessions were genetically different than the breeding lines. A total of 59 single nucleotide polymorphisms (SNPs) associated with seed weight and size traits were discovered across nine chromosomes. Although the harvest window for edamame (for a specific planting date) at R6 is short (<5 – 7 days), the yearly harvest window for edamame at R6 can be from mid-August to early-mid October. The three varieties that are green, black, or brown at

R8 had the best color after pasteurizing. The results of this dissertation will help define a breeding, production, and processing strategy for edamame.

© 2018 by David Moseley
All Rights Reserved

Acknowledgments

I would like to thank my family for your patience and love during this time. I was not able to be around as much as I would want. Hopefully you will always see this time I spent as a sacrifice of love for our family. I also hope I have been and always will be a good example for you.

I would also like to thank Dr. Pengyin Chen for allowing me to have this opportunity, and for lifting me up when I was ready to fall.

I would like to thank my committee members, Dr. Kirkpatrick, Dr. Brye, Dr. Morawicki, and Dr. Ross. I was able to learn from you both in and out of the classroom. I hope I remember to use you as an example as I move forward.

I would like to thank the soybean breeding group for your support during this time. I hope I was able to help you in some way. I especially give thanks to Dr. Mozzoni and Dr. Orazaly for your help on my dissertation. I would also like to thank Dr. Shi, Dr. Kaler, and Dr. Ray for technical assistance for my association mapping study.

Finally, I would like to thank the Arkansas Soybean Promotion Board for the fellowship. It would have been difficult to be successful with this goal without it.

Dedication

To my Lord and Savior Jesus Christ for blessing me well beyond what I deserve. To my family who stood with me during these six years.

Table of Contents

List of Tables

Chapter 2

List of Figures

Chapter 2

Chapter 3

Chapter 1

Introduction and Literature Review

Edamame

Vegetable soybean (edamame) is a food-grade soybean (*Glycine max* (L.) Merill), that is suitable for direct human consumption. The acceptable attributes according to commercial standards are large green pods with gray pubescence preferred (Miles et al., 2000). The pods should be crescent shaped, ≥ 5 cm long, and consist of two to three seed; and the seed should have increased digestibility, firm texture (but not chewy), and sweet flavor (Funatsuki et al., 2006; IDA, 1990; Rackis, 1978; Shanmugasundaram et al., 1989; Watanabe, 1988). Edamame is harvested at the immature (reproductive R6-R7) stage, when the beans fill 80-90% of the pods (Fehr et al., 1971; Konovsky et al., 1994; Shanmugasundaram and Yan, 2004). At maturity, the seed should weigh >30 g/100 seed (Shanmugasundaram and Yan, 2004).

Food Products

The premium edamame product is typically consumed by squeezing the bean straight from the pod after cooking and flavoring to taste; and the pods are discarded as they are not edible (Miles et al., 2000). Edamame can also be shelled and used in soups and salads (Konovsky et al., 1994). After drying at maturity (R8 reproductive growth stage), the seed can either stay green (seed coat alone or seed coat and cotyledon) or turn yellow, black, or brown (Kiuchi et al., 1987; Miles et al., 2000). Dark colored dry edamame is popular in Japan (Miles et al., 2000). Roasted edamame, after maturity, is another product that is becoming popular in the United States (Mentreddy et al., 2002). Preserving high moisture edamame to be shelf stable at room temperature has been attempted within the previous ten years (Czaikoski et al., 2013; Mozzoni et al., 2009) Mozzoni et al. (2009) evaluated methods to sterilize edamame in a brine with high percent hydrogen (pH); however, Czaikoski et al. (2013) investigated methods to pasteurize edamame in an acidic (low pH) brine.

Edamame Market

Edamame, traditionally a Japanese vegetable (Konovsky et al., 1994; Miles et al., 2000), is popular throughout Asian countries (Japan, China, Korea and Taiwan) (Shanmugasundaram

and Yan, 2004); however, the market is gaining momentum in the United States (Konovsky et al., 1994; Miles et al., 2000). In 2012, a Houston, TX based company, constructed what is reported to be the first commercial scale edamame production, processing, and distributing company in the United States: located in Mullberry, Arkansas (Medders, 2012). The growing edamame market in the United States can provide a premium alternative crop for farmers (Konovsky et al., 1994; Harlander, 2002).

Sensory Attributes

Masuda et al. (1989) reported that the most important flavor qualities of edamame are sweet (i.e. sucrose content) and savory (i.e. amino acid content). According to Konovksy et al. (1994), the United States consumer prefers a buttery flavor and texture, while the Japanese consumer prefers a beany flavor. The beany flavor will increase with maturity and is controlled by either linolenic acid or the lipoxygenase enzyme that oxidizes the linolenic acid (Wang et al., 2006). The need of a healthy crop, high in protein and low in fat, creates an opportunity to market edamame. Bullock and Desquilbet (2002) suggested the fact that edamame is not a genetically modified crop makes it more acceptable for direct consumption.

Shanmugasundaram and Yan (2004) reported the standard for seed weight in the Japanese edamame market is >30 grams/100 seed. The United States is thought to not have varieties adapted to local conditions that meet this standard (Shanmugasundaram and Yan, 2004). Currently, the University of Arkansas' advanced edamame germplasm consist of seed that range from 20 – 28 grams/100 seed at maturity; however, there are breeding lines in the pipeline that exceed 30 grams/100 seed.

Health Attributes

The various seed coat colors can have health benefits due to the pigments: the black and brown seed coats accumulate anthocyanins and procyanidins which have been reported to aid in fighting cardiovascular disorder, prevent inflammation, and scavenge harmful radicals (Kim et al., 2006; Nizamutdinova et al., 2009; Takahata et al., 2001).

The current food supply must be doubled in the next 40 years to meet the growing population's demand (Abe et al., 2003; Harlender, 2002). A large part of the world's increasing population consumes an imbalanced diet, leading to malnutrition (Abe et al., 2003). Soybeans are well known for their health benefits (Mebrahtu, 2008), including high amounts of protein (237 ml contains 17 grams of protein), fiber (237 ml contains 8 grams of fiber), calcium, beta carotene, isoflavones (especially genistein and daidzein), and linolenic acid (Alleman et al., 2000; Kris-Etherton et al., 2000; Simonne et al., 2000). They are also low in saturated fats (Meydani et al., 1991). Although grain soybeans have healthy attributes, they also have high amounts of oligosaccharides. These complex carbohydrates cause flatus in human consumers; however, vegetable soybeans have lower amounts of oligosaccharides, resulting in better digestibility (Hymowitz and Collins, 1974). The high protein, fiber, calcium, vitamin A, and antioxidant content (USDA, 2011) in edamame make it a nutritious vegetable with a sweet flavor (Alleman et al., 2000; Miles et al., 2000).

Agronomic practices

Masuda et al. (1989) reported that several agronomic practices can affect flavor including variety selection, planting density, fertilizer application, harvest procedures, and processing conditions.

According to Ashlock (2000b), there are three soybean production systems: early soybean production system (ESPS), full-season production system (FSSPS), and double-crop production system (DCSPS). Each production system has its advantages and disadvantages which are mainly based on soil and ambient temperatures at germination and pod fill, pests, available moisture, maturity group, and day length. The range of planting dates recommended for ESPS are from April 1st to April 30th for south Arkansas and April 7th to May 7th for North Arkansas. The planting date range for FSSPS is between April 25th to June 15th. The planting date range for DCSPS is between June 1st to July 15th. Soybean seed planted from April 15th to June 15th will typically produce maximum yield potential due to several factors (Ashlock et al.,

2000b). The main factor is soybean plants are phototropic, as their reproductive growth stages are triggered by longer hours of darkness (Garner and Allard, 1920) and the day length begins to shorten after the summer solstice in late June. Phototropism is a major factor in yield potential, since the plant will reach maximum yield only if it has produced enough photosynthetic material to produce maximum seed potential. Johnson et al. (1960) indicated that phototropism can affect later stages of reproductive development, not only triggering flowering. Varieties can have a differential sensitivity to delayed planting and phototropism (Johnson et al., 1960); furthermore, very early varieties (i.e. 00 and 0) have been reported to not be sensitive to phototropism (Polson, 1972). In addition, as the relative maturity increases, the soybean reproductive growth stages become increasingly more sensitive to long nights (Johnson et al., 1960; Major et al., 1975).

Ashlock et al. (2000a) recommended planting conventional soybeans 2.54 to 3.81 cm deep at a rate of 33 seeds/m to achieve a final germination rate of 23.76 plants/meter; however, Miles et al. (2000) recommended planting edamame 0.635 – 1.27 cm deep, with a final rate of 13.2 plants/meter. It has been observed that edamame has a lower germination rate at any depth, but is increasingly sensitive as the depth increases (Zhang et al., 2013); therefore, a planting rate close to the suggestion for conventional soybeans (33 seeds / m) should be considered. Caution should be taken when planting edamame shallow, as the seed needs to absorb up to 50% moisture by weight before it germinates (Ashlock et al., 2000a).

Edamame requires adequate phosphorous and potassium input as indicated by a soil test. Typically, nitrogen fertilizer is not recommended for fields previously planted to soybean. This is due to a bacterium, *Bradyrhizobium japonicum* that forms nitrogen-fixing nodules on the root system. If soybean have never been grown in the field, the seeds should be inoculated with *Bradyrhizobium japonicum* (Miles et al., 2000); however, 56 to 112 kg of nitrogen per hectare, for inoculated and not inoculated seed respectively, is recommended for edamame.

Conventional soybean harvest methods cannot be used for edamame, as the pods are stripped off the plants at the R6 reproductive growth stage. Therefore, the commercial industry uses a modified green bean picker to harvest edamame (Miles et al., 2000). Correct harvesting practices are critical to ensure a quality edamame product that has a firm, but not chewy texture (Watanabe, 1988; Wszelaki et al., 2005). A prevalent thought is that the harvest window for premium quality edamame can be as short as 3-4 days (Miles et al., 2000). According to Purcell et al. (2014), the R6 reproductive growth stage, which begins when the seed completely covers the white membrane found inside the pod, can span an average of 18 days. After harvest, the pods should be precooled at 0 – 2.8 ºC to maintain flavor until the pods or shelled beans are blanched and preserved (e.g. quick freeze, freeze dry, or roast) (Tsay and Sheu, 1991).

Sugar content

Kuo et al. (1997) reported the sucrose content of the soybean seed coat drops between the growth stages of R6.2 and R7.0; and of the seed cotyledon between growth stages of R6.2 and R6.5. However, the seed cotyledon accumulates more sucrose between the growth stage of R6.5 and R6.7 (Kuo et al., 1997). Kuo et al. (1997) also stated the seed coat and cotyledon accumulate more stachyose between the growth stages of R6.2 and R7.0; and the raffinose content drops in the seed coat after the R6.2 stage, but rises in the cotyledon. Suarez et al. (1999) stated that soybeans with high concentrations of stachyose and raffinose cause flatulence in humans.

Seed weight heritability and QTL

The standard seed weight for edamame is > 30g/100 seed (Shanmugasundaram and Yan, 2004). Orf et. al. (1999) reported seed weight heritability is 50% and accounted for by many small QTLs. Tinius et al. (1991) supported this heritability report, but added that the heritability for seed weight can be as high as 94 percent. There are over 200 QTLs, across all 20 linkage groups (LG), associated with seed weight listed in soybase (www.soybase.org).

Seed size heritability and QTL

Salas et al. (2006) measured seed size traits as seed height (SH), seed breadth (SB), seed length (SL), and volume (VOL). Volume was calculated as width X height X length. Salas et al. (2006) reports the heritability for all size traits are high (approximately 80-95%). There were 19 QTLs associated with seed size traits distributed over ten (LG). Many of the QTLs explaining seed size are also found by other researchers to explain seed weight (Salas et al., 2006). In addition, Salas et al. (2006) reported transgressive segregation was found for each seed size component, making it possible for breeders to use selective germplasm to breed for desired seed dimensions. Salas et al. (2006) described the measurements as follows: length equals the "longest distance across the seed parallel to the hilum," height equals the "longest distance from top to bottom of the seed," and breadth equals the "longest distance across the seed perpendicular to the hilum." At least 30% or more of the genetic variation for seed shape was explained by a maximum of four QTLs. No epistasis effect between the QTL markers explaining seed size was found by Salas et al. (2006); however, it was noted there may have been an epistasis effect, but was through smaller and undetectable QTLs.

Seed Texture

According to Zhang et al. (2008), seed hardness is a quantitative trait, controlled by multiple genes. Multiple QTL markers explaining seed hardness were found, which could help to screen for texture (Zhang et al., 2008).

Seed Coat Color (Black and Brown)

At maturity, edamame can have green, black, brown, or yellow seed coats and cotyledons (Kiuchi et al., 1987). The various seed coat colors are beneficial to the edamame market due to the health benefits of the pigments. The black and brown seed coats accumulate anthocyanins and procyanidins which have been reported to have several health benefits including fighting cardiovascular disorder, preventing inflammation, and scavenging harmful radicals (Kim et al., 2006; Nizamutdinova et al., 2009; Takahata et al., 2001). The accumulation

of anthocyanins and procyanidins are what make the seed coat black or brown (Todd and Vodkin, 1993). The genetic ability of a soybean to accumulate these pigments are found in the I locus located in LG A2 (Todd and Vodkin, 1993). There are four possible alleles in the I locus, I (dominant- yellow); i^i (pigmented hilum); i^k (regions of saddle-shaped pigmentation), and i (recessive – black or brown seed coat). In addition to the I locus, there are two other loci that control seed coat color; the R locus and T locus. The R locus will determine if the seed coat is black or brown (R=black; r=brown). The T locus will also affect seed-coat color. The T locus has a pleiotropic effect, as it also controls pubescence color. If the dominant T allele is present, a self-colored black (R) soybean will be black. If the soybean is brown (r) with a (T) allele, it will be brown. However, if a self-colored black (R) soybean has a t allele, it will be imperfect black and the brown (r) soybean with a t allele will be buff (Todd and Vodkin, 1993).

Seed Coat and Cotyledon Color (Green)

Approximately half of the 343 edamame type accessions selected for this research from the Germplasm Resource Information Network (GRIN) database, have green seed coats or cotyledons. The cause of a green soybean at maturity is due to a “stay-green” gene (Guiamet et al., 1991). Seed cotyledon color is controlled by two recessive genes, d1 and d2, found in two different linkage groups. The D1 loci is found in the LG D1A and the D2 loci is found in the LG B1. A third locus, G, will cause the seed coat to stay green, but not the cotyledon (Ott et al., 2013). Although the green seed coat and cotyledon is a desirable trait for edamame, the three “stay green” loci also cause the leaves to have delayed senescence, which can cause inability to harvest at maturity (Guiamet et al., 1991).

Number of One, Two, and Three Seed Pods

Commercial standards for edamame require two or three seed pods. The ratio of two or three seed pods to total pods are dependent on genetic (Tischner et al., 2003) and environmental factors (Vega et al., 2001). Tischner et al. (2003) found QTLs controlling pod size (i.e. potential number of seed per pod), and seed set per ovule (i.e. ratio of seeds set and seed

aborted). Potential pod size is highly heritable, controlled by four QTLs (Tischner et al., 2003). The major QTL, explaining 19% of number of ovules per pod, is located on LG F. The four QTLs explaining number of ovules per pod are reported to be linked to loci explaining male sterility, disease resistance, seed weight, and leaflet number (Tischner et al., 2003). Tischner et al. (2003) also reported that seed set per ovule is explained by three QTLs; and are located on LGs M, L, and C1. The QTLs for seed abortion are linked to flowering date, maturity, and water use efficiency (Tischner et al., 2003). Additional stress on the plant due to limitations or excess of water or nutrients, and pest pressure can limit the genetic potential of the crop to set seed and not abort (Board and Tan, 1995; Tischner et al., 2003; Vega et al., 2001).

Immature Pod Weight and Length

Edamame pods harvested at the immature (R6) growth stage were investigated at Virginia State University between 1996-1998 (Mebrahtu and Mohamed, 2006): weight (0.68%) and length (0.81%) were highly heritable traits.

Maturity

Maturity was found to be highly heritable (Orf et al., 1999). Zhang et al. (2004) reported there are 11 QTLs explaining maturity distributed over five linkage groups (A2, B1, C1, I, and M). Linkage group B1 accounted for five of the 11 maturity QTL.

Pod Shattering

As soybean pods mature, they are susceptible to shattering, where the pod opens at the dorsal and ventral sutures (Tsuchiya, 1987). Shattering is a 93% heritable, partially dominant trait, that is controlled by a small number of genes (i.e. 1-3) and has more of an additive than dominant effect (Tsuchiya, 1987). Edamame germplasm tends to express the shattering trait in the United States. The shattering trait is not expressed as prevalent in Japan, due to the cool, humid climate during harvest season (Funatsuki et al., 2008). It has been confirmed there is a major allele associated with shattering, qPDH1, located on LG J between the markers Sat_093 and Sat_366 (located 2.9 cM a part) (Funatsuki et al., 2006; Funatsuki et al., 2008).

Association Mapping

Association mapping (AM) can be used to identify QTL associated to a trait by correlating phenotypic markers and functional single nucleotide polymorphism (SNPs) across germplasm. The ability of AM to identify a SNP within a gene creates higher mapping resolution than linkage mapping (Soto-Cerda and Cloutier, 2012; Zhu et al., 2008).

Objectives

Overall

The overall objective of this Ph.D. dissertation was to discover and test methodologies that can assist the breeding, production, and processing of edamame, which will help the state of Arkansas and her farmers.

Chapter 2: Evaluate Diversity and Association Mapping Of Edamame Germplasm

A total of 343 large - seeded accessions, from seven countries, were selected from the Germplasm Resources Information Network (GRIN) (Perry et al., 1988) to estimate the overall diversity available world-wide. The mapping panel of 343 accessions were analyzed through genome-wide association studies (GWAS) to locate quantitative trait loci (QTLs) associated with seed weight and size traits. The diversity and associated QTLs of the 343 accessions were compared to the University of Arkansas soybean breeding program's large-seeded breeding lines.

Chapter 3: Observe Timing of Edamame Harvest at the R6 Growth Stage

The performance of three commercial edamame varieties were evaluated to assist in defining a planting, maintenance, and harvest schedule. The most optimum harvest time and harvest window to result in the largest collective pod weight and greenest color was observed.

Chapter 4: Preserve Edamame at Room Temperature

Pasteurization techniques were observed to preserve edamame in an acidic brine that is shelf stable at room temperature. The effect of pasteurizing a commercial variety in a brine

consisting of turmeric and sucrose was evaluated. In addition, three varieties from the University of Arkansas were investigated to see if there was a variety effect when pasteurizing edamame.

References

Abe, J., D.H. Xu, Y. Suzuki, A. Kanazawa, and Y. Shimamoto. 2003. Soybean germplasm pools in Asia revealed by nuclear SSRs. Theor. Appl. Genet.106:445–453.

Alleman, G.C. Miles, and T. Lumpkin. 2000. Edamame for nutrition and health. Farming West of the Cascades Series. WA State Univ. Ext. Publ., in press.

Ashlock, L., R. Klerk, G. Huitink, T. Keisling, and E. Vories. 2000a. Planting practices, pp. 1-7. *In* Arkansas Soybean Handbook, MP 197. Ark. Coop. Ext. Ser., Little Rock, AR.

Ashlock, L., W. Mayhew, T. Windham, T. Keisling, R. Klerk, D. Beaty, and G. Lorenz. 2000b. Production systems and economics, pp. 1-7. *In* Arkansas Soybean Handbook, MP 197. Ark. Coop. Ext. Ser., Little Rock, AR.

Board, J.E., and Q. Tan. 2001. Crop ecology, production, management. Crop Sci. 35:846-851.

Bullock, D.S., and M. Desquilbet. 2002. The economics of non-GMO segregation and identity preservation. Food Policy 27:81-99.

Czaikoski, K., R.S. Leite, J.M.G. Mandarino, M.C. Carrao-Panizzi, J.B. da Silva, and E.I. Ida. 2013. Canning of vegetable-type soybean in acidified brine: effect of the addition of sucrose and pasteurization time on color and other characteristics. Ind. Crops and Products. 45:472-476.

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

Funatsuki, H., M. Hajika, S. Hagihara, T. Yamada, Y. Tanaka, H. Tsuji, M. Ishimoto, and K. Fujino. 2008. Confirmation of the location and the effects of a major QTL controlling pod dehiscence, qPDH1, in soybean. Breed. Sci. 58:63-69.

Funatsuki, H., M. Ishimoto, H. Tsuji, K. Kawaguchi, M. Hajika, and K. Fujino. 2006. Simple sequence repeat markers linked to a major QTL controlling pod shattering in soybean. Plant Breed. 125:195-197.

Garner, W.W., and H.A. Allard. 1920. Effect of the relative length of day and night and other factors of the environment on growth and reproduction in plants. J. Agr. Res. 18:553-606.

Guiamet, J.J., E. Schwartz, E. Pichersky, and L.D. Nooden. 1991. Characterization of cytoplasmic and nuclear mutations affecting chlorophyll and chlorophyll-binding proteins during senescence in soybean. Plant Physiol. 96:227-231.

Harlander, S.K. 2002. The evolution of modern agriculture and its future with biotechnology. J. Am. Coll. Nutr. 21:161-165.

Hymowitz, T., and F.I. Collins. 1974. Variability of sugar content in seed of *Glycine max* (L.) Merrill and G. soja Sieb. and Zucc. Agron. J. 66:239-240.

IDA. 1990. Iwate-ken Seikabutsu To Shukka Kikaku Shido Hikkei. Iwate-ken Department of Agriculture, Morioka.

Johnson, H.W., H.A. Borthwick, and R.C. Leffel. 1960. Effects of photoperiod and time of planting on rates of development of the soybean in various stages of the life cycle. Bot. Gaz. 122:77-95.

Kim, H.J., T. Irina, J.M. Park, J. I. Chung, S.C. Shin, and K.C. Change. 2006. Anthocyanins from soybean seed coat inhibit the expression of TNF-α-induced genes associated with ischemia/reperfusion in endothelial cell by NF-κB-dependent pathway and reduce rat myocardial damages incurred by ischemia and reperfusion in vivo. FEBS Letters 580:1391-1397.

Kiuchi, Y., H. Ishikawa, N. Nitta, and T. Sasaki. 1987. Collection of crop genetic resources and their characteristics in Iwate Prefecture I: characteristics of local soybean varieties. Tohoku Nogyo Kenkyu 40:139-40.

Kris-Etherton, P.M., D.S. Taylor, S. Yu-Poth, P. Huth, K. Moriarty, V. Fishell, R.L. Hargrove, G. Zhao, and T. Etherton. 2000. Polyunsaturated fatty acids in the food chain in the United States. AM. J. Clin. Nutr. 71 (suppl): pp. 179 – 188.

Konovsky, J., T.A. Lumpkin, and D. McClary. 1994. Edamame: the vegetable soybean. Pages 173-81 in A.D. O'Rourke (ed), Understanding the Japanese food and agrimarket: a multifaceted opportunity. Haworth Press, Binghamton.

Kuo, T.M., C.A. Lowell, and P.T. Smith. 1997. Changes in soluble carbohydrates and enzymic activities in maturing soybean seed tissues. Plant Sci. 125:1-11.

Major, D.J., D.R. Johnson, J.W. Tanner, and I.C. Anderson. 1975. Effects of daylength and temperature on soybean development. Crop Sci. 15:174-179.

Masuda, R., K. Hashizume, and K. Kaneko. 1989. Effect of holding time before freezing on the constituents and the flavor of frozen green soybeans. Nihon Shokuhin Kogyo Gakkaishi 35:763-70.

Mebrahtu, T. 2008. Analysis of nutritional contents in vegetable soybeans. J. Crop Improv. 21:42.

Mebrahtu, T., and A. Mohamed. 2006. Genetic variation for green pod yield and quality among vegetable soybean genotypes. J. Crop Improv. 16:113-130.

Medders, H. 2012. Vegetable soybean company to make Mulberry 'edamame capital' of U.S. University of Arkansas System. http://arkansasagnews.uark.edu/6583.htm

Mentreddy, S.R., A.I. Mohamed, N. Joshee, and A.K. Yadav. 2002. Edamame: A nutritious vegetable crop. pp. 432-438. In: J. Janick and A. Whipkey (eds.), Trends in new crops and new uses. ASHS Press, Alexandria, VA.

Meydani, S.N., A.H. Lichtenstein, P.J. White, S.H. Goodnight, C.E. Elson, M. Woods, S.L. Gorbach, and E.J. Schaefer. 1991. Food use and health effects of soybean and sunflower oils. J. Am. Coll. Nutr. 10:406-428.

Miles, C.A., T.A. Lumpkin, and L. Zenz. 2000. Edamame. In PNW0525, Pacific Northwest Extension Publication. WA State University.

Mozzoni, L.A., R.O. Morawicki, and P. Chen. 2009. Canning of vegetable soybean: procedures and quality evaluations. Int. J. Food Sci. Technol. 44:1125-1130.

Nizamutdinova, I.T., Y.M. Kim, J.I. Chung, S.C. Shin, Y.K. Jeong, H.G. Seo, et al. 2009. Anthocyanins from black soybean seed coats stimulate wound healing in fibroblasts and keratinocytes and prevent inflammation in endothelial cells. Food Chem. Toxicol. 47:2806-2812.

Orf, J.H., K.Chase, T. Jarvik, L.M. Mansure, P.B. Cregan, F.R. Adler, and K.G. Lark. 1999. Genetics of soybean agronomic traits: I. Comparison of three related recombinant inbred populations. Crop Sci. 39:1642-1651.

Ott, A., Y. Yang, M. Bhattacharyya, H.T. Horner, R. G. Palmer, and D. Sandhu. 2013. Molecular mapping of D1, D2, and ms5 revealed linkage between the cotyledon color locus D2 and the male-sterile locus ms5 in soybean. Plants 2:441-454

Perry, M., A.K. Stoner, and J.D. Mowder. 1988. Plant germplasm information management system: Germplasm Resources Information Network. HortSci. 23:56–60.

Polson, D.E. 1972. Day-neutrality in soybeans. Crop Sci. 12:773-776.

Purcell, L.C., M. Salmeron, and L. Ashlock. 2014. Soybean growth and development, pp. 1-8. *In* Arkansas Soybean Handbook, MP 197. Ark. Coop. Ext. Ser., Little Rock, AR.

Rackis, J.J. 1978. Biochemical changes in soybeans: maturation, post-harvest storage and processing, and germination. pp. 34-76. *In* H.O. Hultin and M. Milner (eds), Post-harvest Biology and Technology. Food and Nutr., Westport.

Salas, P., J.C. Oyarzo-Llaipen, D. Wang, K. Chase, and L. Mansur. 2006. Genetic mapping of seed shape in three populations of recombinant inbred lines of soybean (*Glycine max L. Merr.*). Theor. Appl. Genet. 113:1459-1466.

Shanmugasundaram, S., and M.R. Yan. 2004. Global expansion of high value vegetable soybean. VII world soybean conference/IV international soybean processing and utilization conference/III Brazilian Soybean Congress, 29 Feb–5 Mar 2004. Foz do Iguassu, Brazil, pp. 915–920.

Shanmugasundaram, S., T.C.S. Tsou, and S.H. Cheng. 1989. Vegetable soybean in the East. In: ed. A.J. Pascale, World Soybean Research Conference IV Proceedings. Asociacion Argentina de la Soja, Buenos Aires, Argentina, pp. 979-1986.

Simonne, A.H., M. Smith, D.B. Weaver, T. Vail, S. Barnes, and C.I. Wei. 2000. Retention and changes of soy isoflavones and carotenoids in immature soybean seeds (edamame) during processing. J. Agric. Food Chem. 48:6061-6069.

Soto-Cerda, B.J., and S. Cloutier. 2012. Association mapping in plant genomes. In: Caliskan M (ed) Genetic diversity in plants. In Tech, Rijeka. pp. 29-54.

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

Takahata, Y., M. Ohnishi-Kameyama, S. Furuta, M. Takahashi, and I. Suda. 2001. Highly polymerized procyanidins in brown soybean seed coat with a high radical-scavenging activity. J. Agric. Food Chem. 49:5843-5847

Tinnius, C.N., J.W. Burton, and T.E. Carter, Jr. 1991. Recurrent selection for seed size in soybean: I. response to selection in replicate populations. Crop Sci. 31:1137-1141.

Tischner, T., L. Allphin, K. Chase, J.H. Org, and K.G. Lark. 2003. Crop breeding, genetics and cytology. Crop Sci. 43:464-473.

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

Tsay, L.M., and S.C. Sheu. 1991. Studies on the effects of cold storage and precooling on the quality of vegetable soybeans. In: S. Shanmugasundaram (Ed), Vegetable soybean: research needs for production and quality improvement. Asian Vegetable Research and Development Center. pp. 113-119.

Tsuchiya, T. 1987. Physiological and genetic analysis of pod shattering in soybeans. JARQ. 21:166-175.

USDA. 2011. Department of agriculture of the united states. National nutrition database 205 for standard reference release 26. Search from: http://www.nal.usda.gov/fnic/foodcomp/search/. Accessed on: 26 May, 2014.

Vega, C.R.C., F.H. Andrade, V.O. Sadras, S.A. Uhart, and O.R. Valentinuz. 2001. Seed number as a function of growth. A comparative study in soybean, sunflower, and maize. Crop Sci. 41:748-754.

Wang, L., R. Guan, L. Zhangxiong, R. Chang, and L. Qiu. 2006. Genetic diversity of Chinese cultivated soybean revealed by SSR markers. Crop Sci. 46:1032-1038.

Watanabe, M. 1988. Tunneru Saibai – Sapporo-midori. Nogyo Gijutsu Taikei (yasai*)* 13:1-8.

Wszelaki, A.L., J.F. Delwiche, S.D. Walker, R.E. Liggett, S.A. Miller, and M.D. Kleinhenz. 2005. Consumer liking and descriptive analysis of six varieties of organically grown edamame-type soybean. Food Qual. Pref. 16:651-658.

Zhang, B., P. Chen, C.Y. Chen, D. Wang, A. Shi, A. Hou, and T. Ishibashi. 2008. Qantitative trait loci mapping of seed hardness in soybean. Crop Sci. 48:1341-1349.

Zhang, Q.Y., M. Hashemi, S.J. Hebert, and Y.S. Li. 2013. Different responses of pre-emergence and early seedling growth to planting depth between vegetable soybean and grain soybeans*.* Legume Res. 36:515-521.

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 (*Glycine max* L. Merr.) genetic map and their association with EST markers. Theor. Appl. Genet. 108:1131-1139.

Zhu, C., M. Gore, E.S. Buckler, and J. Yu. 2008. Status and prospects of association mapping in plants. Plant Genome 1:5-20.

Chapter 2

Evaluation of Genetic Diversity and Association Mapping for Seed Weight and Size in Vegetable Soybean (*Glycine max* L. Merr.) Germplasm

Abstract

Vegetable soybean (edamame, *Glycine max* (L.) Merr) is a food-grade soybean that is either harvested at the R6 reproductive stage, while still green, or at maturity for products such as roasted edamame. A seed weight of >30g/100 seed is one of the most important characteristics of marketable edamame. Therefore, increasing seed weight of U.S. edamame breeding lines is a necessary objective in developing new varieties and understanding diversity among large-seeded germplasm is critical to define breeding strategies. The objectives of this study are: i) compare the diversity between the University of Arkansas' breeding lines with the overall edamame-type germplasm available through the Germplasm Resources Information Network (GRIN), and ii) identify new and previously reported quantitative trait loci (QTL) associated with seed weight (SW) and volume (VOL), along with their components: seed length (SL), seed breadth (SB), and seed height (SH) through genome-wide association studies (GWAS) using single nucleotide polymorphisms (SNPs) from the SoySNP50K iSelect BeadChip. A total of 343 accessions from seven countries were ordered from GRIN using a search criterion of >20g/100 seed. In addition, 31 breeding lines from the University of Arkansas were analyzed using a 6K SNP chip. The accessions were planted in a randomized complete block with two replications at Fayetteville and Stuttgart, AR in 2014 and 2015. There were two main genetic groups among the 343 accessions: one consisting of Japanese lines and another of South Korean lines. The accessions were genetically dissimilar when compared to the 31 breeding lines from the University of Arkansas, with the breeding lines in the top half of the phylogeny tree and the accessions in the bottom half. After analyzing with best linear unbiased prediction (BLUP) values across all environments and deleting all non-significant SNPs in at least three out of four environments, 47, 38, 9, 8, and 61 SNPs were observed to be associated with SW, VOL, SL, SB, and SH, respectively. Several SNPs were associated with more than one trait; ss715609782 located at the 2.62 Mb position on chromosome 11 was associated with

four traits (SW, VOL, SH and SB). Thirty-one and 24 SNPs were associated with SW, VOL, and SH on chromosome 4 and 20, respectively. The SNP (ss715587475) positioned at the 24.89 Mb position on chromosome 4 was located within the gene *Glyma.04G143300*, which has been associated with seed weight in *Arabidopsis*. Eight of the SNPs associated with seed weight identified by the association panel were available in the 6K SNP chip. Breeding value estimation of the 31 Arkansas breeding lines using these markers suggests a positive trend of association. This confirms the QTLs reported in the association mapping are also present in the current breeding germplasm. This study will help edamame breeding efforts by identifying valuable germplasm sources and key molecular markers to target in marker assisted breeding efforts.

Introduction

Vegetable soybean (*Glycine max* (L.) Merr.) are a food-grade soybean commonly referred to as edamame, or “branched bean” (Jian, 1984). Edamame is more desirable for human consumption than conventional soybean due to increased digestibility, sweeter flavor, and larger seed size (Rackis, 1978). The premium edamame product is harvested immature at the R6 reproductive stage while the pods are green with no blemishes and the pods should be crescent shaped with two or three seed (Sirisomboon et al., 2007). Gray pubescents is more desirable than tawny color due to aesthetics, as the edamame is often consumed from the pod as a snack food (Mentreddy et al., 2002).

In addition to the immature R6 reproductive stage, edamame can be harvested at the mature (R8) reproductive stage. The seed weight (SW) of edamame at the mature (R8) stage should be >30g/100 seed, compared to conventional soybeans weighing approximately 14-16g/100 seed (Shanmugasundaram and Yan, 2004). When mature and dry, the seed color will be either yellow, black, brown, or green due to remaining pigments when the chlorophyll breaks down The pigments can be beneficial as they contribute healthy attributes such as antioxidants (Kim et al., 2006; Nizamutdinova et al., 2009; Takahata et al., 2001).

According to Nuss (2013), between 22,600 to 27,000 Mg of edamame are consumed annually in the United States, however, between 70% (Nolen et al., 2016) to 95% (Ogles et al., 2016) of the edamame consumed in the US is imported from China and Taiwan. Since 2012, production and processing of edamame has increased in the US; however, additional breeding and research is still required for the US to develop varieties with larger seed size and better adaptability to be more competitive in the market. (Ogles et al., 2016; Nuss, 2016).

Genetic diversity in a breeding program is a critical component to develop improved varieties (Shi et al., 2010.) East Asia is thought to be the center of domestication for soybeans; therefore, germplasm from Asia should be used in any breeding effort (Dong et al., 2013). Abe et al., (2003) reported there are two main germplasm pools from Japanese and Chinese soybean populations. Different Korean germplasm were found to be grouped with either Japanese or Chinese accessions. Zhang et al. (2010a) observed large-seeded cultivars and germplasm collected from six states within the United States had less diversity than genotypes from South Korea and Japan.

Seed weight is an important characteristic for edamame production. Orf et. al. (1999) reported seed weight heritability is 50% accounted for by many small QTLs. Tinius et al. (1991) supports the heritability report by Orf et al. (1999), but added the heritability for SW can be as high as 94%.

The results from Zhang et al. (2016) suggested that many small effect loci control seed weight. Zhang et. al. (2004) found four QTLs explaining SW distributed over three chromosomes, 8, 11, and 17. As a result of genome wide association studies (GWAS), Lara (2016) found 16 single nucleotide polymorphisms (SNPs) associated with SW located on ten chromosomes and Yan et al. (2017) found eight SNPs on two chromosomes.

Seed size is an important trait in vegetable soybean as it is a component of the weight and has important aesthetic properties (Liang et al., 2005). Seed size traits are composed of seed length (SL), seed breadth (SB), and seed height (SH) (Hu et al., 2013; Salas et al., 2006).

The product of SL X SB X SH make up the seed volume (VOL) (Salas et al., 2006). Salas et al. (2006) describes SL, SB, and SH as the longest distances “parallel to the hilum”, “perpendicular to the hilum”, and from “top to bottom” (where the hilum marks the top), respectively.

Salas et al. (2006) noted seed size traits were inherited quantitatively and the progeny of bi-parental populations had a normal distribution with transgressive segregation. The heritability of seed size traits has been estimated as 0.58-0.97, 0.42-0.98, 0.72-0.98, and 0.44-0.88 for SL, SB, SH, and VOL, respectively (Hu et al., 2013; Salas et al., 2006). Salas et al. (2006) reported all seed size traits (except SL*SB) had a strong correlation with a r^2 value between 0.5-0.9 and the correlation between SL and SB was either not significant or low (0.3) across three populations. However, seed size traits are not correlated with the overall seed shape (from long and thin to round). A lack of correlation makes it possible to breed for a desired shape, either long or thin or round, while maintaining a large seed weight (Cober et al., 1997; Salas et al., 2006).

The objectives of this research project were to: i. compare the phenotypic and genotypic diversity among 343 diverse, large-seeded accessions from GRIN and 31 breeding lines from the University of Arkansas breeding program, ii. identify SNPs in proximity to previously reported and new quantitative trait loci (QTL) associated with SW, SL, SB, SH, and VOL., and iii. calculate breeding values (BV) using SNPs associated with seed weight. It was hypothesized that: i. the University of Arkansas large-seeded breeding lines were genetically separate from the GRIN accessions, conventional, and small-seeded soybean., ii. multiple QTLs control SW and seed size traits., and iii. favorable alleles associated with SW will increase the BV of the line.

Materials and Methods

Plant Material

A total of 343 accessions from seven countries were selected from the Germplasm Resources Information Network (GRIN) (Perry et al., 1988) to analyze the available diversity

and conduct association mapping for seed weight and size traits. The accessions were selected by searching the GRIN database for seed >20 g/100 seed at the R8 reproductive stage. The maturity groups (MG) of the accessions consisted of 000 to 9, with 82% belonging to the MGs III – VII. To compare the diversity available world-wide to the University of Arkansas large-seeded breeding lines, 29 accessions from GRIN and 31 Arkansas breeding lines were selected. Of the 31 Arkansas breeding lines, 26 were large-seeded (~18-32 g/100 seed), three were conventional (~13-16 g/100 seed), and two were small-seeded (~10-12 g/100 seed).

Field Experiments

The 343 accessions were planted at two locations in two consecutive years (2014 and 2015) for a total of four environments (ENV): Fayetteville 2014 (14FAY), Fayetteville 2015 (15FAY), Stuttgart 2014 (14STU), and Stuttgart 2015 (15STU). In 2014, the seed were planted in July; however, in 2015 the seed were planted in May. The experimental design was a randomized complete block (RCB) with two replications. The Fayetteville and Stuttgart locations were the Arkansas Agricultural Research and Extension Center in Fayetteville and the Rice Research and Extension Center in Stuttgart, respectively. The soil at these sites were Leaf silt loam (fine, mixed, active, thermic Typic Albaquults) in Fayetteville and Dewitt silt loam (fine, smectitic, thermic Typic Albaqualfs) in Stuttgart (Soil Survey Staff, 2017). The plots in Stuttgart followed rice production (both years) and the plots in Fayetteville followed fallow ground (both years). In both locations (2014 and 2015), the fields were cultivated before planting to ensure a uniform soil bed. There were 33 seeds planted per row and the rows were 3 m long and 0.91 m wide (FAY) and 3 m long and 0.76 m wide (Stuttgart). Approximately 250 grams of pods, per entry, were randomly harvested at maturity (R8 reproductive stage).

Potassium and phosphorus fertilizer were applied to the experimental plots with rates suggested by soil test from the University of Arkansas. The plots were irrigated as needed based on visual observation of the soil moisture content. The plots in Stuttgart were irrigated twice in 2014 and five times in 2015; and the plots in Fayetteville were irrigated 5 times in 2014

and 2015. Weeds were controlled by applying Select Max, Flexstar, and Storm in Stuttgart in 2014; Valor, Select, and Basagran in Stuttgart in 2015; and Charger Max, Scepter 70DG, and Flexstar in Fayetteville in 2014 and 2015. Stink bugs (*Nezara viridula*) were controlled in Fayetteville in 2014 and 2015 by applying Grizzly Z.

Phenotyping

The seed weight (g/100 seed) was determined by collecting 100 random seed and weighing with a precise scale sensitive to the hundredth of a gram (Zhang et al., 2016). Seed length, SB, SH, and VOL were measured with the protocol demonstrated by Nelson and Wang (1989) and repeated by Hu et al. (2013), Niu et al. (2013), Salas et al. (2006), and Xu et al. (2011), where 20 seed were randomly selected and measured with a vernier caliper sensitive to the thousandth of a millimeter. The volume was determined by calculating the product of the SL, SB, and SH measurements as performed by Salas et al. (2006).

Statistical Analysis

The procedures PROC means, PROC univariate, and PROC Corr (α=0.05) of SAS 9.4 (SAS Institute, 2014) were used to determine descriptive statistics, normality, and Pearson correlations, respectively. The PROC varcomp procedure of SAS 9.4 (SAS Institute, 2014), with the restricted maximum likelihood (REML) method, was used to calculate the broad sense heritability (H^2), on an entry-mean basis, of each trait.

To calculate H^2, the genotype nested within MG, ENV, and the interaction of genotype nested within MG by ENV were considered as random variables. As described by Kaler et al. (2017), the formula for H^2 was:

$$H^2 = \sigma^2_G/(\sigma^2_G + (\sigma^2_{Genv}/l) + (\sigma^2_\epsilon/lb))$$

where variances were due to: σ^2_G [genotype(MG)], σ^2_{Genv} [genotype(MG) by ENV], and σ^2_ϵ (error); and l and b were the number of ENV and blocks, respectively.

Best linear unbiased prediction (BLUP) values across all ENV were calculated by the PROC glimmix procedure of SAS 9.4 (SAS Institute, 2014) to minimize the ENV variation. The

factors genotype nested within MG, ENV, block nested within ENV, and the interaction of genotype nested within MG by ENV were considered as random effects. The least square means (LSM) for each ENV (14FAY, 14STU, 15FAY, 15STU) were calculated by the PROC mixed procedure of SAS 9.4 (SAS Institute, 2014) with the method = type 3 option and the kenwardroger adjustment. The factors for the LSM values were genotype nested within MG (fixed), and block (random). For the GWAS analysis, BLUP values were used across all environments (AAE), and the LSM values were used for each individual environment.

Genotyping

A total of 42,509 single nucleotide polymorphisms (SNPs) from The SoySNP50K iSelect BeadChip (assembly version Wm82.a1) were downloaded from Soybase (www.soybase.org), for the association mapping panel of 343 accessions. After deleting SNPs located at unanchored sequenced scaffolds, consisting of a minor allele frequency (MAF) of ≤5%, and missing or heterogenous SNPs ≥ 2%, 22,272 SNPs were utilized for the GWAS analysis. To compare the diversity between the 343 accessions to the University of Arkansas breeding lines Deoxyribonucleic acid (DNA) was extracted from the 60 genotypes as described by (Lara, 2016), which was sent to the University of Minnesota Genomics Center for genotyping using a 6K SNP chip.

The mixed linear model (MLM) method from The Genome Association Prediction Integrated Tool (GAPIT) R package (Zhang et al. 2010b) was used to estimate the linkage disequilibrium (LD) by the squared correlation (r^2) of allele frequency (Weir and Cockerham, 1996).

Population Structure and Diversity

The population structure of the 343 accessions and the University of Arkansas breeding lines were analyzed through an admixture model using the program STRUCTURE 2.3.4 (Pritchard et al., 2000). The burn-in period was set to 10,000 with 20,000 Monte Carlo Markov Chain (MCMC) replicates. The number of clusters (K) was set to 1-10 with 10 iterations for each

cluster. The optimum number of K was calculated using the Evanno criterion (Evanno et al., 2005) method and STRUCTURE Harvester (Earl, 2012) was used to estimate the best fitting number of clusters.

The 343 accessions and 31 breeding lines were assigned to a Q group based on the data from STRUCTURE 2.3.4. The results from STRUCTURE Harvester (Earl, 2012) suggested the 343 accessions should be in two groups. Therefore, the cutoff to assign an accession to either Q1 or Q2 was 0.55. The results from STRUCTURE Harvester set the optimum K for the 60 lines consisting of 31 Arkansas breeding lines and 29 accessions to four; therefore, the cutoff to assign a genotype to a Q group within the 60 lines was 0.5. If a line did not have a value of at least 0.55 and 0.5 for 343 accessions and the 60 genotypes, respectively, the lines with the largest values were added together as an admixture until a value of 0.55 or 0.5 was achieved.

The diversity of the 343 accessions and the 31 Arkansas breeding lines was illustrated by phylogeny trees using the maximum likelihood tree method from Mega 7 (Kumar et al., 2016). The parameters for Mega 7 in this study were the same as specified by Shi et al. (2016).

Genome Wide Association Mapping

A MLM using principal component analysis (PCA = 5) and kinship (K) to account for population structure (Zhao et al., 2007) and family relatedness (Yu et al., 2006), respectively, was analyzed using the Tassel (Bradbury et al., 2007) and GAPIT (Zhang et al., 2010b) software. In addition, a generalized linear model (GLM) using a Q matrix from Structure (Pritchard et al., 2000) was analyzed through Tassel.

A SNP was considered associated with SW or a seed size trait if the $-Log_{10}(p)$ value was >2.5. Two steps were took to find associated SNPs. First, SNPs associated using the BLUP values across all ENV were selected. Second, SNPs associated with the traits in each individual ENV were analyzed using least square means. For this research, a SNP was suggested to be associated with a trait if the $–Log_{10}(p)$ value was >2.5 AAE and in at least three out of four

environments. A threshold of three out of four ENV is stricter than the majority of the reviewed literature.

Manhattan plots and gBlup plots used in this research were from GAPIT (Zhang et al., 2010b). The allelic effect for each SNP associated with a trait of interest was calculated by Tassel (Bradbury et al., 2007). The BV of each accession was determined by the summation of all allelic effects (favorable and unfavorable) for each accession (Kaler et al., 2017).

Results

Phenotypic Variation

All traits were observed to be normally distributed according to an Anderson-Darling A-Sq goodness of fit test ($p<0.005$). The range for SW, SL, SB, SH, and VOL were 36.18 g/100 seed, 7.8 mm, 5.26 mm, 4.1 mm, and 565.98 mm^3, respectively (Table 1).

The traits SW, SL, SH, and VOL had a strong correlation between them, with SH correlated with VOL and SW with a coefficient of 0.871 and 0.828, respectively. Seed breadth had a low correlation with SW and SH with a coefficient of 0.409 and 0.244, respectively. Seed breadth had a negative correlation with seed length (-0.368). All correlations were significant ($p<0.001$) (Table 2).

There were differences in all seed weight and size traits observed among the accessions ($p<0.0001$). The broad sense heritability (H^2) was high (0.89-0.95) for SW, SL, SB, SH, and VOL, where seed length was the most heritable (0.95) (Table 4).

The ENV effect was significant ($p<0.05$) for all traits except for SH ($p=0.052$) (Table 3). The 10 top and bottom ranked accessions for SW, SL, SB, SH, and VOL are shown in Tables 5, 6, 7, 8, and 9, respectively. The largest mean for SW (29.68 mm), SB (6.33 mm), and VOL (459.15) was at Stuttgart in 2014. The largest mean for SL (9.11 mm) was in Fayetteville in 2015; however, the lowest mean for SB (5.95 mm) was also in Fayetteville in 2015. The lowest mean for SW (26.38 mm), SL (8.61 mm), and VOL (412.36 mm^3) was in Fayetteville in 2014 (Tables

5-9). Phenotypic data (BLUP across all environments), MG, and country of origin for all 343 entries are listed in (Supplementary table 1).

Genetic Variation

The peak of the delta K from STRUCTURE 2.3.4 software was K = 2, (Figure 1) indicating there were two main Q groups between the mapping panel of 343 accessions from GRIN. Groups 1 and 2 consisted predominately of South Korean and Japan accessions, respectively, with a few accessions from other countries split between the two groups. Two and one lines from South Korea and China, respectively, were an admixture between the two Q groups. The phylogenetic tree of the 343 accessions support the Q grouping with the accessions from South Korea and Japan on opposite sides (Figure 2). Accessions from North Korea and China were located in both groups; however, accessions from the USA were located in the Japanese group (Figure 2).

The peak of the delta K for the group of 60 lines, consisting of 31 University of Arkansas breeding lines and 29 accessions was K = 4 (Figure 1). The place of origin for Q groups one, two, three, and four were predominantly: Japan, University of Arkansas, University of Arkansas, and Korea (North and South), respectively. The results of the phylogenetic tree indicate the breeding lines from the University of Arkansas belong in one half, and the accessions in another half. The first five breeding lines were both conventional and small-seeded releases from the University of Arkansas. The next 28 lines consisted of one accession from Japan and 27 large-seeded breeding lines. The last 27 lines including one breeding line and 26 accessions were from other countries (Figure 3).

QTL Discovery

There were 22,272 SNPs used to perform the GWAS with the mapping panel consisting of 343 accessions. The SNPs were located across all 20 chromosomes, which have an approximate total length of 950 Mb and an average of 47.5 Mb (Table 10) per chromosome. The average SNP distance across all 20 chromosomes was 45.08 kb with a range of 26.14 kb

(chromosome 13) to 67.13 kb (chromosome 1) (Table 10). There was an average of 24 SNPs per Mb with a range of 15 SNPs/Mb (chromosomes 1 and 20) and 38 SNPs/Mb (chromosome 13) (Table 10).

Twenty percent of the SNPs had a minor allele frequency between 0.05-0.10. Thirteen percent of alleles had a MAF between 0.11-0.15. The remaining groups (grouped by MAF of 0.05) had a MAF between 8-10 percent (Table 11). The r^2 value across all chromosomes dropped to 0.25 at a genetic distance of approximately 250 kbp (Figure 4). This rate of LD decay is similar to the LD decline reported by Kaler et al. (2017).

The GWAS results, using BLUP values AAE, show there were 73-140 SNPs associated with the seed weight and size traits (Figure 5). When filtering for SNPs associated in at least three of the four environments, the number of associated SNPs were: 47, 38, 9, 5, and 61, for SW, VOL, SL, SB, and SH, respectively (Figure 6).

The SNPs were located on five, four, six, two and five chromosomes for SW, VOL, SL, SB, and SH, respectively (Figure 6). There were two new SNPs detected for SL on chromosome 5 with significance levels of 3.02 and 2.75; and one new SNP detected for SL on chromosome 19 with a significance level of 2.72 (Table 12; Figure 6). There were four new SNPs, within 1 Mb, detected for SB on chromosome 19 with significance levels between 2.57-2.95 (Table 13 and Figure 6).

A total of 49 SNPs across five chromosomes were associated with more than one trait. (Figure 7). Chromosome 4 had 31 SNPs, between the 19.11 Mb and 33.59 Mb position, associated with SH (Table 14; Figure 6) and 22 and 11 of them were also associated with SW (Table 15; Figure 6) and VOL (Table 16; Figure 6), respectively. The SNP located at the 24.89 Mb position (ss715587475) was associated AAE where SW, VOL, and SH had a significance level of 5.29, 4.25, and 5.4, respectively; and was associated with each trait in all four environments. The allele effect for ss715587475 was 4.39 g/100 seed, 59 mm^3, and 0.51 mm for SW, VOL, and SH, respectively; and the r^2 value was 0.063, 0.049, and 0.065 for SW, VOL,

and SH, respectively. ss715609782 located at the 2.62 Mb position on chromosome 11 was associated with SW, VOL, SH, and SB (Tables 13-16; Figure 7) with a significance level of 4.29, 4.51, 3.29, and 4.06, respectively. Another SNP, ss715612171 located at the 32.47 Mb position on chromosome 12, was associated with SH and SL (Tables 12, 14; Figure 7) with a significance level of 3.10 and 3.09, respectively. The SNP ss715630059, located at the 26.28 Mb position on chromosome 18, was associated with SW, VOL, and SH with a significance level of 3.62, 3.26, and 4.13 (Tables 14-16; Figure 7), respectively. A total of 24 SNPs between the 24.38 and 26.79 Mb position on chromosome 20 was associated with VOL and SH, in which 22 were also associated with SW (Tables 14, 16; Figure 7). The SNP located at the 26.33 Mb position (ss715637113) was associated with SW, VOL, and SH with a significance level of 4.35, 3.87, and 4.27, respectively (Tables 14-16). The effect for SW, VOL, and SH was 3.33 g/100 seed, 47.57 mm^3, and 0.38 mm, respectively. The r^2 value for SW, VOL, and SH was 0.050, 0.044, and 0.049, respectively.

Breeding Values

The effect of the major allele on the phenotypic trait was calculated using Tassel software (Bradbury et al., 2007). The traits with all major alleles of associated SNPs having a positive effect on the phenotype were SW, SB and VOL (Tables 13, 15, 16). For SL, five major alleles had a positive effect and 4 major alleles had a negative effect (Table 12). There were 60 major alleles with a positive effect on SH; however, one major allele had a negative effect (Table 14). The average allelic effect for the traits were: SW (3.20 g/100 seed), SB (0.28 mm), VOL (46.26 mm3), SL (0.40 or -0.59 mm), and SH (0.37 or -0.20 mm) (Tables 12-16). The breeding values (BV) for each accession were calculated by the summation of all associated alleles for each trait. The accessions with the top and bottom 10 values for each trait were identified according to the gRank value (Tables 5-9). Seventeen accessions were in the top ten for multiple traits. The six accessions that had more than one trait among SL, SB, and SH were: PI416876 (SL, SH), PI417322 (SL, SH), PI506556 (SB, SH), PI506606 (SL, SH), PI506744 (SL,

SH), and PI506746 (SL, SH). The remaining six accessions were in either the top ten for SB or SH paired with SW or volume. There were 10 accessions in the bottom ten for multiple traits. Among the 10 accessions, there were four ranking as the lowest for SL and SH (PI424574, PI506697, 507038, PI593979); all four were also in the lowest ten for SW and volume. Four of the remaining six accessions were ranked in the lowest for SW, SH, and volume (PI408228B, PI423909, PI445847, PI96783). The last two were in the bottom ten for SW and VOL (PI194647) or SH and VOL (PI417436). For SW, SB, SH, and VOL, the accession with the largest and lowest gRank value also had the largest and lowest pRank value respectively. The same accession also had the highest gRank and pRank for seed length. The exception was for the smallest gRank value for SL; however, the three accessions with the lowest gRank value also had the three lowest pRank values.

The remaining accessions in the top ten for each trait had similar gRank and pRank values with a few exceptions. Overall, the accessions with high or low BV scores also had high or low pRank values, respectively. The highest and lowest BV for SW was 150 and -150, respectively. The same trend was observed for SH and VOL.

A total of eight of the 47 SNPs associated with SW were found in the 6K SNP chip from the University of Minnesota. Within the 31 University of Arkansas breeding lines and 29 GRIN accessions, 15 had all eight favorable alleles, where 6 had all unfavorable alleles. The two group's average seed size were 25.24 g/100 seed and 17.62 g/100 seed, respectively (Table 17). Two breeding lines, R07-10396 and V96-7198, had all eight favorable SW alleles. The genotypes that had only one or two favorable SW alleles generally had a lower seed weight. The two SW QTL commonly found in the University of Arkansas breeding program is located on chromosomes 12 and 20 (Figure 6).

Discussion

Phenotypic and Genetic Diversity

This research evaluated and compared the phenotypic and genetic diversity of seed weight and seed size traits among 343 large-seeded accessions from seven different countries to 31 breeding lines from the University of Arkansas soybean breeding program. Two previous research projects were observed to have investigated the relationship between seed weight and size traits (Kato et al., 2014; Xie et al., 2014). However, Xie et al. (2014) conducted QTL mapping for SW and size traits, where Kato et al. (2014) mapped QTL for only seed weight.

Ten previous research projects were observed to have genotypes within the mapping population weighing at least 20g/100 seed (Han et al., 2012; Kato, 2014; Kim et al., 2010; Lara, 2016; Maughan et al., 1996; Sun et al., 2012; Teng et al., 2009; Xie et al., 2014; Yan et al., 2017; Zhang et al., 2016); however, only four of them mapped QTLs to genotypes weighing over 30 g/100 seed (Kato, 2014; Lara, 2016; Sun et al., 2012; Yan et al., 2017). Of the four studies observed to conduct mapping of QTLs to seed length, two had a range of approximately 6-10 mm (Salas et al., 2006; Xu et al., 2011), and two had a range of approximately 5 -13.5 mm (Hu et al., 2013; Niu et al., 2013). Likewise, the range for seed height examined by Salas et al. (2006), Xu et al. (2011), and Niu et al. (2013) was approximately 4 – 7.5 mm, where Hu et al. (2013) had a seed height range of 4.39 – 9.54 millimeters. The SB range for Hu et al. (2013), Salas et al. (2006), and Xu et al. (2011) was approximately 3 – 8 mm, where Niu et al. (2013) had a SB range of 4.14 – 9.5 millimeters.

The only previous studies mapping QTLs to seed weight with a range close to what was investigated in this research (10 - 46.18 g/100 seed) were Lara (2016) and Yan et al. (2017). This research also had a maximum SL value of 14.30 mm and minimum SB value of 3.6 mm. A SL value of 14.33 mm, was longer than what was observed in the literature. Hu et al. (2013) had a minimum SB value of 2.81 mm, which was the only SB value lower than was examined in this research. However, the SB range for Hu et al. (2013) was 2.81 – 6.48 mm, where current

research had a range of 3.6 – 8.86 mm. It is possible the phenotypic diversity in this population increased the ability to map associated QTLs to seed weight and size traits (McCarthy et al., 2008).

Salas et al. (2006) reported there was a strong positive correlation between all seed size traits, except between SB and SL, which had a low correlation. The results of this research also suggest a strong positive correlation between SW, VOL, SH, and seed length. Seed height was correlated with SB and VOL with a coefficient of 0.828 and 0.871, respectively. However, the results of this research suggest a low correlation between SB and SH (coefficient of 0.244) and a negative correlation between SB and SL with a coefficient of -0.368.

The broad sense heritability was high for all seed weight and size traits with a range of 0.89 to 0.95, which agrees with the results reviewed in the literature (Cober et al., 1997; Salas et al., 2006, Yan et al., 2017.) The ability to find QTLs associated with the seed weight and size traits AAE and in at least three out of four ENV was possible due to the high heritability.

The results from STRUCTURE Harvester (Earl, 2012) in this research suggest there are two main groups (Japanese and South Korean) among the 343 accessions; with other countries split between the two groups. These results are contrary to Abe et al. (2003), who reported the two main Asian groups were of Japanese and Chinese origin. However, one reason for this difference is 45% and 35% of our accessions came from Japan and South Korea respectively, where 5% came from China.

The University of Arkansas large-seeded breeding lines were genetically different than the accessions, conventional, and small-seeded breeding lines. This agrees with Zhang et al. (2010a), who suggested the large-seeded germplasm in the United States had less diversity than Asian germplasm. There was one accession from Japan grouped with the large-seeded breeding lines. Therefore, South Korean lines may be more genetically diverse to the University of Arkansas breeding lines than the Japanese lines. Since the large-seeded germplasm seems to have narrow diversity and has been shown to be genetically different than soybeans from

Asian countries, selecting parents from Asia should be considered in edamame breeding programs (Dong et al., 2013).

QTL Discovery

The number of chromosomes in which SNPs were associated for each trait were: five for SW, four for VOL, six for SL, two for SB, and five for seed height. There were 35 SNPs across four chromosomes that were associated with SH, SW, and VOL (Figure 6). This makes sense due to the strong correlations between the three traits. Two and one new SNPs on chromosomes 5 and 19, respectively, were detected for SL (Figure 6). There were four new SNPs, within 1 Mb, associated with SB on chromosome 19 (Figure 6). It is possible that the diversity of SB and SL analyzed in this research increased the ability to find new QTLs associated with the traits.

There were a total of 49 SNPs associated with more than one trait, where 47 of them were associated with a combination of SH with SW and/or VOL on chromosomes 4, 18, and 20 (Figure 7). Chromosome 11 had a SNP associated with SH and SB, and chromosome 12 had a SNP associated with SH and SL (Figure 7). The association of a SNP to two different size traits indicates a pleiotropic effect. Although these SNPs were in regions where SW or VOL QTL are located in soybase (http://soybase.org) and in the literature (Lara, 2016; Yan et al., 2017), the SNPs detected on chromosomes 4, 18, and 20 are new for SH and VOL. In addition, the SNP observed on chromosome 11 is new for SH and seed breadth. Finally, the SNP noted on chromosome 12 is new for SH and SL. Forty-eight of the 49 SNPs had a significance level >3.0 (Tables 12-16).

On chromosome 4, between the 19.1 and 33.6 Mb position, there were 31, 22, and 11 SNPs associated with SH, SW, and VOL, respectively (Figure 7). Although 30 of the 31 SNPs had a significance level for each trait at >3.0, the SNP at the 24.89 Mb position, ss715587475, was associated AAE for SH, SW, and VOL, with a significance level of 5.4, 5.29, and 4.25, respectively (Tables 14-16). The SNP type is C/A; and the C allele is the major and favorable

allele. This SNP explained 6.5% of the variation for SH and had an allele effect of 0.51 mm. For SW, this SNP explained 6.3% of the variation, with an allele effect of 4.39 g/100 seed. For VOL, this SNP explained 4.9% of the variation and had an allele effect of 59 mm^3. This SNP is within seed weight markers SW47-3, SW20-2, SW36-15, and SW45-3 (Grant et al., 2010), and is in between two seed weight SNPs (located at the 24.27 and 27.91 Mb positions), as documented by Yan et al. (2017). In addition, this SNP was within a seed size QTL discovered by Salas et al. (2006). Yan et al. (2017) noted the gene *Glyma.04G143300* was in chromosome 4, near the SW SNPs they discovered. It was also discussed that this gene has been associated with seed weight in *Arabidopsis* (Yan et al., 2017). The SNP ss715587475 at the 24,888,097 bp position, observed to be associated with SW and seed size traits in this research, is within the gene *Glyma.04G143300*.

On chromosome 20, between the 24.4 and 26.8 Mb positions, there were 24 SNPs associated with SH and VOL and 22 SNPs associated with SW. Although all 24 SNPs were associated with a significance level of >3.0, the SNP ss715637113, located at the 26.33 Mb position, was associated with SW, SH, and VOL with a significance level of 4.35, 4.27, and 3.87, respectively, and was associated with SW and SH in all four environments. This SNP explained 5% of the variation for SW and had an allele effect of 3.33 g/100 seed. For SH, this SNP explained 4.9% of the variation and had an allele effect of 0.38 mm. Finally, this SNP explained 4% of the variation for VOL and had an allele effect of 47.57 mm^3. This SNP, located at the 26.33 Mb position, is within the chromosome bp position of SW markers: SW34-5, SW35-5, and SW9-1; and is in between two SNPs (26.09 and 26.50 Mb) Lara (2016) suggested were associated with SW. The significance level and bp position of this QTL indicate it may be a causative SNP for seed weight and seed size traits.

Breeding Values

The BV of the accessions agreed with (Kaler et al., 2017); where the BV were similar to the phenotypic rankings (Tables 5-9). The accessions with all 47 favorable SW alleles and all 47

alternative alleles had a BV of 150 and -150, respectively, with a positive trend in between. The same extremes were observed for SH and VOL. The high and low BV for SL was 3.1 and -1.8 respectively; and the high and low BV for SB was 2.3 and -1.7, respectively.

The same association between favorable alleles and phenotypic value was observed in the 60 genotypes consisting of 31 breeding lines from the University of Arkansas and 21 accessions from GRIN. Of the 47 SNPs associated with SW, eight were also in the 6K SNP chip. A total of 15 genotypes had all eight favorable alleles with a SW of 25.24 g/100 seed; and six genotypes had zero favorable alleles with a SW average of 17.62 g/100 seed (Table 17). The eight SNPs associated with SW in the 31 breeding lines are located on three chromosomes. A total of six, one, and one SNPs are located on chromosomes 4, 12, and 20, respectively. The alleles on chromosomes 4 and 20 are within 612 kb to the SNPs associated with SW (with the highest level of significance) observed in the 50K SNP chip. Two breeding lines, R07-10396 and V96-7198 had all eight favorable alleles, suggesting each had at least three out of five SW loci detected in the mapping of 343 accessions. The SNPs associated with SW observed on chromosomes 11 and 18 were not available in the 6K SNP chip; therefore, no data were available for the Arkansas breeding lines. The SNPs associated with SW on chromosomes 12 and 20 were common throughout the University of Arkansas large-seeded breeding lines. These results indicate the QTLs associated with SW on chromosomes 4, 12, and 20 are present in the Arkansas edamame breeding program; however, the QTL located on chromosome 4 was not as common. Breeding efforts to combine the QTLs on chromosomes 4, 12, and 20 should continue. In addition, accessions containing the favorable alleles for QTLs associated with SW located on chromosomes 11 and 18 should be used in breeding efforts to increase the University of Arkansas' edamame lines to >30g/100 seed.

Conclusions

There were two main genetic groups (Japanese and South Korean origin) among the 343 GRIN accessions. The University of Arkansas large-seeded breeding lines were genetically

different from the accessions, and from the conventional and small-seeded breeding lines. Compared to the University of Arkansas breeding lines, the accessions from South Korea appear to be more diverse than from Japan.

A total of 59 SNPs associated with SW, SL, SH, and VOL were observed across seven chromosomes (4, 7, 11, 12, 16, 18, 20). There were two new SNPs associated with SL on chromosomes 5 and 19; and four new SNPs, within 1 Mb, associated with SB on chromosome 19. This research discovered SNPs associated with SW on five chromosomes: 4, 11, 12, 18, and 20. This data suggest the SNPs on chromosomes 4, 12, and 20 are present among the University of Arkansas breeding lines, with the SNPs on chromosomes 12 and 20 more prevalent.

The validity of allele effect of the SNPs associated with SW were reinforced by calculating BV for each accession and breeding line. The accessions and breeding lines that had the highest and lowest seed weight generally had the largest and lowest BV, respectively. This data can assist breeders in selecting parents and progeny that have multiple seed weight and size QTLs, resulting in improved edamame varieties.

References

Abe, J., D.H. Xu, Y. Suzuki, A. Kanazawa, and Y. Shimamoto. 2003. Soybean germplasm pools in Asia revealed by nuclear SSRs. Theor. Appl. Genet. 106:445-453.

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

Cober, E.R., H.D. Voldeng, and J.A. Frégeau-Reid. 1997. Heritability of seed shape and seed size in soybean. Crop Sci. 37:1767-1769.

Dong, D., X. Fu, F. Yuan, P. Chen, S. Zhu, B. Li, Q. Yang, X. Yu, and D. Zhu. 2013. Genetic diversity and population structure of vegetable soybean (*Glycine max* (L.) Merr.) *In* China as revealed by SSR markers).

Earl, D.A. 2012. STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method. Conser. Genet. Resour. 4:359-361.

Evanno, G., S. Regnaut, and J. Goudet. 2005. Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. Mol. Ecol. 14:2611-2620. Doi:10.1111/j.1365-294X.2005.02553.x.

Grant, D., R.T. Nelson, S.B. Cannon, and R.C. Shoemaker. 2010. SoyBase, the USDA-ARS soybean genetics and genomics database. Nucl. Acids Res. 38 (suppl 1): D843-D846. doi: 10.1093/nar/gkp798.

Han, Y., D. Li, D. Zhu, H. Li, X. Li, W. Teng, and W. Li. 2012. QTL analysis of soybean seed weight across multi-genetic backgrounds and environments. Theor. Appl. Genet. 125:671-683.

Hu, Z., H. Zhang, G. Kan, M. Deyuan, D. Zhang, G. Shi, et al. 2013. Determination of the genetic architecture of seed size and shape via linkage and association analysis in soybean (*Glycine max* L. Merr.). Genetica 141:247-254.

Jian, Y. 1984. Situation of soybean production and research in China. Tropical Agriculture Res. Ser. 17:67-72.

Kaler, A.S., J.D. Ray, W.T. Schapaugh, C.A. King, and L.C. Purcell. 2017. Genome-wide association mapping of canopy wilting in diverse soybean genotypes. Theor. Appl. Genet. 130:2203-2217.

Kato, S., T. Sayama, K. Fujii, S. Yumoto, Y. Kono, Tae-Young, Hwang, et al. 2014. A major and stable QTL associated with seed weight in soybean across multiple environments and genetic backgrounds. Theor. Appl. Genet. 127:1365-1374.

Kim, E.H., S.H. Kim, J.I. Chung, H.Y. Chi, J.A. Kim, and I.M. Chung. 2006. Analysis of phenolic compounds and isoflavones in soybean seeds (*Glycine max* (L.) Merrill) and sprouts grown under different conditions. Eur. Food Res. Technol. 222:201 https://doi.org/10.1007/s00217-005-0153-4.

Kim, H.K., Y.C. Kim, S.T. Kim, B.G. Son, Y.W. Choi, J.S. Kang, et al. 2010. Analysis of quantitative loci (QTLs) for seed weight and fatty acid composition using recombinant inbred lines in soybean. J. Life Sci. 20:1186-1192.

Kumar, S., G. Stecher, and K. Tamura. 2016. MEGA7: molecular evolutionary genetics analysis version 7.0 for bigger datasets. Mol. Biol. Evolution. 33:1870-1874.

Lara, S.L.M. 2016. Association mapping of seed weight, protein, and sucrose content; and kinetics of edamame under infrared treatment. Theses and Dissertations. 1747. http://scholarworks.uark.edu/etd/1747.

Liang, H.Z., W.D. Li, H. Wang, and X.J. Fang. 2005. Genetic effects on seed traits in soybean. Acta. Genet. Sin. 32:1199-1204.

Maughan, P.J., M.A. Saghai Maroof, and G.R. Buss. 1996. Molecular-marker analysis of seed-weight: genomic locations, gene action, and evidence for orthologous evolution among three legume species. Theor. Appl. Genet. 93:574-579.

Mentreddy, S.R., A.I. Mohamed, N. Joshee, and A.K. Yadav. 2002. Edamame: A nutritious vegetable crop. pp. 432-438. *In*: J. Janick and A. Whipkey (eds.), Trends in new crops and new uses. ASHS Press, Alexandria, VA.

McCarthy, M.I., G.R. Abecasis, L.R. Cardon, D.B. Goldstein, J. Little, J.P.A. Ioannidis, and J.N. Hirschhorn. 2008. Genome-wide association studies for complex traits: consensus, uncertainty and challenges. Nature Rev. Genet. 9:356-369.

Nelson, R.L., and P. Wang. 1989. Variation and evaluation of seed shape in soybean. Crop Sci. 29:147-150.

Niu, Y., Y. Xu, X.F. Liu, S.X. Yang, S.P. Wei, F.T. Xie, and Y.M Zhang. 2013. Association mapping for seed size and shape traits in soybean cultivars. Mol. Breed. 31:785-794.

Nizamutdinova, I.T., Y.M. Kim, J.I. Chung, S.C. Shin, Y.K. Jeong, H.G. Seo, et al. 2009. Anthocyanins from black soybean seed coats stimulate wound healing in fibroblasts and keratinocytes and prevent inflammation in endothelial cells. Food Chem. Toxicol. 47: 2806-2812.

Nolen, S., B. Zhang, and M.K. Kering. 2016. Increasing fresh edamame bean supply through season extension techniques. J. Hortic. 3:170. doi:10.4172/2376-0354.1000170.

Nuss, J. 2013. US soybean farmers see growth potential in edamame. CNSnews.com. http://www.cnsnews.com/news/article/us-soybean-farmers-see-growth-potential-edamame (accessed 11 Jan. 2017).

Ogles, C.Z., E.A. Guertral, and D.B. Weaver. 2016. Edamame cultivar evaluation in central Alabama. Agron. J. 108:2371-2378.

Orf, J.H., K.Chase, T. Jarvik, L.M. Mansure, P.B. Cregan, F.R. Adler, and K.G. Lark. 1999. Genetics of soybean agronomic traits: I. comparison of three related recombinant inbred populations. Crop Sci. 39:1642-1651.

Perry, M., A.K. Stoner, and J.D. Mowder. 1988. Plant germplasm information management system: Germplasm Resources Information Network. HortSci. 23:56–60.

Pritchard, J.K., M. Stephens, and P. Donnelly. 2000. Inference of population structure using multilocus genotype data. Genetics 155:945-959.

Rackis, J.J. 1978. Biochemical changes in soybeans: maturation, post-harvest storage and processing, and germination. Pages 34-76. *In*: H.O. Hultin and M. Milner *(eds),* Post-harvest Biology and Technology. Food and Nutrition, Westport.

Salas, P., J.C. Oyarzo-Llaipen, D. Wang, K. Chase, and L. Mansur. 2006. Genetic mapping of seed shape in three populations of recombinant inbred lines of soybean (*Glycine max L. Merr.*). Theor. Appl. Genet. 113:1459-1466.

SAS Institute. 2014. The SAS System for Windows. Version 9.4. SAS Inst. Inc., Cary, NC.

Shanmugasundaram, S., and M.R. Yan. 2004. Global expansion of high value vegetable soybean. VII world soybean conference/IV international soybean processing and utilization conference/III Brazilian Soybean Congress, 29 Feb–5 Mar 2004. Foz do Iguassu, Brazil. 915–920.

Shi, A., B. Buckley, B. Mou, D. Motes, J.B. Morris, J. Ma, H. Xiong, J. Qin, W. Yang, J. Chitwood, Y. Weng, and W. Lu. 2016. Association analysis of cowpea bacterial blight resistance in USDA cowpea germplasm. Euphytica 208:143-155.

Shi, A., P. Chen, B. Zhang, and A. Hou. 2010. Genetic diversity and association analysis of protein and oil content in food-grade soybeans from Asia and the United States. Plant Breed. 129:250-256.

Sirisomboon, P., P. Pornchaloempong, and T. Romphophak. 2007. Physical properties of green soybean: Criteria for sorting. J. Food Eng. 79:18-22.

Soil Survey Staff, Natural Resources Conservation Service, United States Department of Agriculture. Web Soil Survey. Available online at the following link: https://websoilsurvey.sc.egov.usda.gov/. Accessed [12/11/2017].

Sun, Y.N., J.B. Pan, X.L. Shi, X.Y. Du, Q. Wu, Z.M. Qi et al. 2012. Multi-environment mapping and meta-analysis of 100 seed weight in soybean. Mol. Biol. Rep. 39:9435-9443.

Takahata, Y., M. Ohnishi-Kameyama, S. Furuta, M. Takahashi, and A. Suda. 2001. Highly polymerized procyanidins in brown soybean seed coat with a high radical-scavenging activity. J. Agric. Food Chem. 49:5843-5847.

Teng, W., Y. Han, Y. Du, D. Sun, Z. Zhang, L. Qiu, et al. 2009. QTL analysis of seed weight during the development of soybean (*Glycine max* L. Merr.) Heredity 102:372-380.

Tinnius, C.N., J.W. Burton, and T.E. Carter, Jr. 1991. Recurrent selection for seed size in soybean: I. response to selection in replicate populations. Crop Sci. 31:1137-1141.

Weir, B.S., and C. Cockerham. 1996. Genetic data analysis II: Methods for discrete population genetic data. Sinauer Assoc. Inc. Sunderland, MA. USA.

Xie, F.T., Y. Niu, J. Zhang, S.H. Bu, H.Z. Zhang, Q.C. Geng, J.Y. Feng, and Y.M Zhang. 2014. Fine mapping of quantitative trait loci for seed size traits in soybean. Mol. Breed. 34:2165-2178.

Xu, Y., H.N. Li, G.J. Li, X. Wang, L.G. Cheng, and Y.M. Zhang. 2011. Mapping quantitative trait loci for seed size traits in soybean (*Glycine max* L. Merr.) Theor. Appl. Genet. 122:581-594.

Yan, L., N. Hofmann, S. Li, M.E. Ferreira, B. Song, G. Jiang et al. 2017. Identification of QTL with large effect on seed weight in a selective population of soybean with genome-wide association and fixation index analyses. BMC Genomics 18:529.

Yu, J., G. Pressoir, W.H. Briggs, B.I. Vroh, M. Yamasaki, J.F. Doebley et al. 2006. A unified mixed-model method for association mapping that accounts for multiple levels of relatedness. Nat. Genet. 38:203-208.

Zhang, B., P. Chen, S.L. Florez-Palacios, A. Shi, A. Hou, and T. Ishibashi. 2010a. Seed quality attributes of food-grade soybeans from the U.S. and Asia. Euphytica 173:387-396.

Zhang, Z., E. Ersoz, C.Q. Lai, R.J. Todhunter, H.K. Tiwari, M.A. Gore, et al. 2010b. Mixed linear model approach adapted for genome-wide association studies. Nat. Genet. 42:355-360.

Zhang, J., Q. Song, P.B. Cregan, and G.L. Jiang. 2016. Genome-wide association study, genomic prediction and marker-assisted selection for seed weight in soybean (*Glycine max*). Theor. Appl. Genet. 129:117-130.

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

Zhao, K., M.J. Aranzana, S. Kim, C. Lister, C. Shindo, C. Tang, et al. 2007. An *Arabidopsis* example of association mapping in structured samples. PLoS Genet 3: e4. https://doi.org/10.1371/journal.pgen.0030004.

Table 1. Descriptive statistics of seed weight and size traits.

Trait	Mean	SD[a]	Range	Skewness	Kurtosis
SW (mm)	27.61	4.93	10-46.18	-0.03	0.46
SL (mm)	8.92	0.95	6.5-14.3	1.37	3.13
SB (mm)	6.11	0.67	3.6-8.86	-0.66	0.92
SH (mm)	7.9	0.54	5.73-9.83	-0.48	1.02
VOL (mm^3)	431.07	73.55	166.74-732.72	-0.10	0.50

Mean, SD, Skewness, and Kurtosis are from across all environments
Range is from all four environments
Test for normality: Anderson-Darling A-Sq: Pr > A-Sq <0.005
[a] Standard deviation

Table 2. Correlation coefficients (r) between seed weight and size traits.

	SH[e]	SL[d]	SB[c]	VOL[b]	SW[a]
SH	1.000				
SL	0.615	1.000			
SB	0.244	-0.368	1.000		
VOL	0.871	0.537	0.550	1.000	
SW	0.828	0.566	0.409	0.900	1.000

*For all correlations ($p<0.001$)
[a] Seed weight (g/100 seed)
[b] Volume (SH X SL X SB) (mm^3)
[c] Seed breadth
[d] Seed Length
[e] Seed height

Table 3. Analysis of variance (ANOVA) table for seed weight and size traits across all environments.

Trait	ANOVA (mean squares)					
	MG[e]	ent(MG)[d]	ENV[c]	block(ENV)[b]	ENV*ent(MG)[a]	Residual
SW (g/100s)	1474.1	106.2	1864.2	162	8.4	4.2
SL (mm)	37.5	4.8	34.9	2.4	0.1	0.1
SB (mm)	10.9	2.1	17.8	2	0.1	0
SH (mm)	17.3	1.6	6.8	1.1	0	0
VOL (mm^3)	239795	18265	284908	35064	2242.9	1124.7

P<0.001 for all variables except ENV. The P values for ENV were p=0.02, p=0.01, p=0.03, p=0.052, and p=0.03 for SW, SL, SB, SH, and VOL respectively.

[a] Interaction of environment by genotype nested within maturity group

[b] Block nested within environment

[c] Environment

[d] Genotype nested within maturity group

[e] Maturity group

Table 4. Variance components and broad sense heritability (H^2) for seed weight, length, breadth, height, and volume.

Trait	G[d]	G*ENV[c]	ϵ[b]	H^{2a}
SW (g/100s)	14.73	4.99	5.08	0.89
SL (mm)	0.71	0.1	0.12	0.95
SB (mm)	0.33	0.05	0.07	0.94
SH (mm)	0.21	0.04	0.05	0.93
VOL (mm^3)	3257.2	990.36	1266.20	0.89

[a] Broad sense heritability
[b] Residual
[c] Genetic by environment variation
[d] Genetic variation

Table 5. The top 10 and bottom 10 ranked accessions (gRank) for seed weight based on the gBLUP values. Also shown is the pRank value based on the average seed weight across all four environments (AAE), and the individual seed weight value from each environment: Fayetteville 2014/2015 (14FAY, 15FAY) and Stuttgart 2014/2015 (14STU, 15STU).

Accession	Seed Weight (g/100 seed)					pRank[d]	gBLUP[c]	gRank[b]	BV[a]
	14FAY	14STU	15FAY	15STU	AAE[e]				
Largest SW									
PI 506990	33	46	39	43	39	1	36	1	**150**
PI 181564	34	40	34	38	36	5	34	2	**150**
PI 417322	33	36	35	37	35	9	34	3	**147**
PI 506744	34	.	34	.	35	12	34	4	**147**
PI 506752	33	34	36	37	35	11	34	5	**147**
PI 506556	30	39	34	43	36	6	34	6	**150**
PI 416876	33	.	40	.	36	4	34	7	**150**
PI 506579	30	.	36	.	33	25	34	8	**150**
PI 506746	37	34	37	42	37	2	34	9	**147**
PI 506606	33	35	35	36	35	13	34	10	**147**
Smallest SW									
PI 194647	17	21	21	19	20	334	22	334	**147**
PI 96783	19	21	18	21	20	331	21	335	**-17**
PI 445847	19	22	17	20	20	333	21	336	**15**
PI 423909	17	20	16	18	18	338	21	337	**147**
PI 507038	19	20	19	18	19	336	21	338	**-150**
PI 227213	16	23	.	.	20	332	21	339	**-17**
PI 424574	18	17	17	18	18	339	20	340	**-147**
PI 408228B	17	18	15	17	17	341	20	341	**-124**
PI 506697	14	15	12	14	14	342	16	342	**-150**
PI 593979	13	15	12	13	14	343	16	343	**-150**

[a] Breeding value (summation of allelic effects of all associated SNPs with the trait of interest)
[b] Genetic rank (calculated by GAPIT)
[c] Genetic Blup values (calculated by GAPIT)
[d] Phenotypic ranking
[e] Phenotypic value across all environments

Table 6. The top 10 and bottom 10 ranked accessions (gRank) for seed length based on the gBLUP values. Also shown is the pRank value based on the average seed length across all four environments (AAE), and the individual seed length value from each environment: Fayetteville 2014/2015 (14FAY, 15FAY) and Stuttgart 2014/2015 (14STU, 15STU).

Accession	Seed Length (mm)					pRank[d]	gBLUP[c]	gRank[b]	BV[a]
	14FAY	14STU	15FAY	15STU	AAE[e]				
Largest SL									
PI 416876	13	.	14	.	13	1	12.2	1	3.1
PI 506800A	12	12	13	12	12	3	11.9	2	3.1
PI 506800B	12	13	13	13	12	4	11.8	3	3.1
PI 506799	11	12	12	12	12	6	11.8	4	3.1
PI 417322	11	11	12	12	12	7	11.6	5	0.5
PI 506746	11	11	13	12	12	5	11.6	6	0.5
PI 506606	11	11	12	12	12	8	11.6	7	0.5
PI 417099	12	.	13	13	13	2	11.4	8	0.5
PI 506996	9	.	.	.	10	40	11.3	9	0.5
PI 506744	11	.	12	.	11	10	11.2	10	1.8
Smallest SL									
PI 549069	7	8	8	9	8	327	7.9	334	-1.3
PI 538408	7	.	7	.	7	340	7.9	335	-1.3
PI 561236	8	8	8	8	8	329	7.9	336	-2.9
PI 561234	8	8	9	8	8	321	7.9	337	-2.9
PI 561241	7	8	8	8	8	332	7.9	338	-2.9
PI 506754	.	.	7	.	8	336	7.9	339	-1.3
PI 507038	8	8	8	7	8	334	7.9	340	-1.3
PI 424574	7	7	7	7	7	343	7.5	341	-2.6
PI 506697	7	7	7	7	7	342	7.5	342	-0.9
PI 593979	7	7	7	7	7	341	7.5	343	-1.8

[a] Breeding value (summation of allelic effects of all associated SNPs with the trait of interest)
[b] Genetic rank (calculated by GAPIT)
[c] Genetic Blup values (calculated by GAPIT)
[d] Phenotypic ranking
[e] Phenotypic value across all environments

Table 7. The top 10 and bottom 10 ranked accessions (gRank) for seed breadth based on the gBLUP values. Also shown is the pRank value based on the average seed breadth across all four environments (AAE), and the individual seed breadth value from each environment: Fayetteville 2014/2015 (14FAY, 15FAY) and Stuttgart 2014/2015 (14STU, 15STU).

Accession	Seed Breadth (mm)					pRank[d]	gBLUP[c]	gRank[b]	BV[a]
	14FAY	14STU	15FAY	15STU	AAE[e]				
Largest SB									
PI 506903	7	7	7	8	7	1	7.0	1	2.3
PI 458141	7	8	7	7	7	3	6.9	2	2.3
PI 417233	7	7	7	7	7	4	6.9	3	2.3
PI 200544	7	7	7	7	7	5	6.8	4	2.3
PI 507031	7	7	6	7	7	26	6.8	5	1.7
PI 243551	7	.	7	.	7	9	6.8	6	1.7
PI 507179	7	7	7	7	7	7	6.8	7	2.3
PI 4243371	7	.	7	.	7	6	6.8	8	2.3
PI 417270	6	8	7	7	7	18	6.8	9	2.3
PI 506556	7	8	7	7	7	17	6.8	10	2.3
Smallest SB									
PI 578470	4	5	4	4	4	339	4.8	334	1.7
PI 506801B	5	5	4	5	5	334	4.7	335	2.3
PI 416928	4	5	4	5	5	336	4.7	336	-1.7
PI 506996	5	.	.	.	5	325	4.7	337	-1.7
PI 417099	4	.	4	4	5	337	4.7	338	-1.7
PI 416876	4	.	4	.	4	340	4.6	339	1.6
PI 506801A	5	5	4	4	4	338	4.5	340	-1.7
PI 506800B	4	4	4	4	4	342	4.3	341	-1.7
PI 506799	5	4	4	4	4	341	4.3	342	-1.7
PI 506800A	4	4	4	4	4	343	4.3	343	-1.7

[a] Breeding value (summation of allelic effects of all associated SNPs with the trait of interest)
[b] Genetic rank (calculated by GAPIT)
[c] Genetic Blup values (calculated by GAPIT)
[d] Phenotypic ranking
[e] Phenotypic value across all environments

Table 8. The top 10 and bottom 10 ranked accessions (gRank) for seed height based on the gBLUP values. Also shown is the pRank value based on the average seed height across all four environments (AAE), and the individual seed height value from each environment: Fayetteville 2014/2015 (14FAY, 15FAY) and Stuttgart 2014/2015 (14STU, 15STU).

Accession	Seed Height (mm)					pRank[d]	gBLUP[c]	gRank[b]	BV[a]
	14FAY	14STU	15FAY	15STU	AAE[e]				
Largest SH									
PI 416876	9	.	10	.	9	1	9.0	1	22.6
PI 417322	9	9	9	9	9	5	8.9	2	22.6
PI 506744	9	.	9	.	9	3	8.9	3	22.6
PI 506746	9	9	10	9	9	2	8.9	4	22.6
PI 506606	9	9	9	9	9	4	8.9	5	22.6
PI 506752	9	9	9	9	9	7	8.8	6	22.6
PI 548457	9	9	9	8	8	30	8.7	7	22.6
PI 417270	8	9	9	9	9	16	8.5	8	22.6
PI 506556	8	9	9	9	9	15	8.5	9	22.6
PI 506579	9	.	9	.	9	13	8.5	10	22.3
Smallest SH									
PI 417436	7	7	7	7	7	330	7.3	334	6.5
FC 199762	7	8	7	7	7	331	7.2	335	-5.6
PI 507038	7	7	7	6	7	332	7.2	336	-22.3
PI 96783	7	7	6	7	7	337	7.0	337	-5.6
PI 424574	7	7	7	7	7	338	7.0	338	-21.4
PI 423909	7	7	7	6	7	339	6.9	339	22.2
PI 408228B	7	7	6	6	7	340	6.9	340	-19.7
PI 445847	7	7	6	6	6	341	6.7	341	5.6
PI 506697	6	6	6	6	6	342	6.5	342	-22.6
PI 593979	6	7	6	6	6	343	6.4	343	-22.6

[a] Breeding value (summation of allelic effects of all associated SNPs with the trait of interest)
[b] Genetic rank (calculated by GAPIT)
[c] Genetic Blup values (calculated by GAPIT)
[d] Phenotypic ranking
[e] Phenotypic value across all environments

Table 9. The top 10 and bottom 10 ranked accessions (gRank) for seed volume based on the gBLUP values. Also shown is the pRank value based on the average seed volume across all four environments (AAE), and the individual seed volume value from each environment: Fayetteville 2014/2015 (14FAY, 15FAY) and Stuttgart 2014/2015 (14STU, 15STU).

Accession	Seed Volume (mm^3)					pRank[d]	gBLUP[c]	gRank[b]	BV[a]
	14FAY	14STU	15FAY	15STU	AAE[e]				
Highest VOL									
PI 506990	524	665	592	614	589	1	548	1	1758
PI 417270	444	616	523	574	535	12	530	2	1758
PI 506556	459	599	536	589	537	10	530	3	1758
PI 506579	497	.	650	.	558	6	530	4	1758
PI 181564	520	622	557	578	564	4	529	5	1758
PI 200544	529	637	582	583	579	2	528	6	1665
PI 417322	504	567	554	551	538	9	513	7	1758
PI 506570	483	.	516	511	508	28	511	8	1758
PI 506744	499	.	549	.	524	15	509	9	1758
PI 408058	493	519	487	505	499	42	507	10	1758
Lowest VOL									
PI 96783	302	332	298	340	322	328	342	334	523
PI 194647	300	316	336	315	321	330	342	335	1758
PI 417436	326	283	287	276	304	336	342	336	-284
PI 507038	329	328	300	262	307	334	333	337	-1758
PI 423909	266	290	290	270	285	339	332	338	1758
PI 445847	306	330	245	296	301	337	325	339	-429
PI 424574	253	278	274	280	277	340	319	340	-1758
PI 408228B	261	287	246	244	266	341	311	341	-1649
PI 506697	229	232	183	208	221	342	248	342	-1758
PI 593979	210	242	191	200	219	343	248	343	-1758

[a] Breeding value (summation of allelic effects of all associated SNPs with the trait of interest)
[b] Genetic rank (calculated by GAPIT)
[c] Genetic Blup values (calculated by GAPIT)
[d] Phenotypic ranking
[e] Phenotypic value across all environments

Table 10. The chromosome sequence length (bp), number of SNPs analyzed within each chromosome, the average distance between each SNP (kb), and the number of SNPs found within a million base pairs (Mb).

Chr	Chr length (bp)	SNPs	Marker distance (kb)	SNP density (#SNPs/Mb)
1	55,915,595	833	67.13	15
2	51,656,713	1376	37.54	27
3	47,781,076	832	57.43	17
4	49,243,852	1001	49.19	20
5	41,936,504	794	52.82	19
6	50,722,821	1137	44.61	22
7	44,683,157	1204	37.11	27
8	46,995,532	1457	32.25	32
9	46,843,750	820	57.13	18
10	50,969,635	1121	45.47	22
11	39,172,790	1000	39.17	26
12	40,113,140	767	52.30	19
13	44,408,971	1699	26.14	38
14	49,711,204	1072	46.37	22
15	50,939,160	1475	34.54	29
16	37,397,385	1045	35.79	28
17	41,906,774	1058	39.61	25
18	62,308,140	1776	35.08	29
19	50,589,441	1091	46.37	22
20	46,773,167	714	65.51	15

Table 11. Distribution of 22,272 single nucleotide polymorphism (SNP) markers grouped by minor allele frequency (MAF).

MAF[b]	Number of SNPs[a]	Percentage (%)
0.05-0.10	4361	20
0.11-0.15	2982	13
0.16-0.20	2282	10
0.21-0.25	2276	10
0.26-0.30	2262	10
0.31-0.35	2239	10
0.36-0.40	1949	9
0.41-0.45	2090	9
0.46-0.50	1831	8

[a] Number of single nucleotide polymorphisms
[b] Minor allele frequency

Table 12. The SNPs associated with seed length across all environments and in at least three of the four environments using GAPIT and Tassel software with a threshold of $-\log_{10}(p) \geq 2.5$.

SL SNP ID	Chr	Position	Allele	MAF[d]	Effect[c]	r^2	$-\log_{10}(p)$[b]	ENV[a]
ss715591833	5	39,666,580	G/A	0.32	0.43	0.031	3.02	14F/15F/15S
ss715591835	5	39,669,594	C/T	0.31	0.40	0.027	2.75	14F/15F/15S
ss715596074	7	13,700,399	C/T	0.09	-0.62	0.050	4.56	14F/14S/15F/15S
ss715612171	12	32,474,069	G/A	0.23	0.27	0.032	3.09	14S/15F/15S
ss715623850	16	26,106,217	G/A	0.093	0.45	0.032	3.11	14S/15F/15S
ss715624487	16	31,837,545	T/C	0.11	-0.49	0.040	3.75	14F/14S/15F/15S
ss715624488	16	31,840,819	A/G	0.06	-0.82	0.058	5.18	14F/14S/15F
ss715632847	18	8,995,306	T/G	0.3	0.46	0.049	4.44	14F/14S/15F/15S
ss715633048	19	11,010	G/A	0.14	-0.44	0.027	2.72	14S/15F/15S

[a] The individual environments the SNP is found to be associated with seed length
[b] Significance level of SNPs in the across all environments GWAS analysis
[c] Allele effect
[d] Minor allele frequency

Table 13. The SNPs associated with seed breadth across all environments and in at least three of the four environments using GAPIT and Tassel software with a threshold of $-\log_{10}(p) \geq 2.5$.

SB SNP ID	Chr	Position	Allele	MAF[d]	Effect[c]	r^2	$-\log_{10}(p)$[b]	ENV[a]
ss715609782	11	2,622,133	G/T	0.24	0.29	0.044	4.06	14F/15F/15S
ss715636273	19	7,386,845	A/G	0.14	0.34	0.03	2.95	14S/15F/15S
ss715636367	19	8,267,096	C/T	0.15	0.3	0.027	2.71	14S/15F/15S
ss715636369	19	8,286,237	G/A	0.12	0.31	0.026	2.57	14S/15F/15S
ss715636375	19	8,309,440	G/A	0.15	0.31	0.028	2.77	14S/15F/15S

[a] The individual environments the SNP is found to be associated with seed breadth
[b] Significance level of SNPs in the across all environments GWAS analysis
[c] Allele effect
[d] Minor allele frequency

Table 14. The SNPs associated with seed height across all environments and in at least three of the four environments using GAPIT and Tassel software with a threshold of $-\log_{10}(P) \geq 2.5$.

SH SNP ID	Chr	Position	Allele	MAF[d]	Effect[c]	r^2	$-\log_{10}(p)$[b]	ENV[a]
ss715587306	4	19,107,515	A/G	0.10	0.37	0.034	3.12	14F/15F/15S
ss715587326	4	19,903,602	A/G	0.09	0.4	0.040	3.59	14F/15F/15S
ss715587338	4	20,426,164	T/C	0.08	0.45	0.047	4.13	14F/14S/15F/15S
ss715587357	4	20,796,034	A/G	0.09	0.37	0.036	3.26	14F/15F/15S
ss715587360	4	20,856,983	T/C	0.09	0.4	0.040	3.59	14F/15F/15S
ss715587370	4	21,252,334	A/G	0.09	0.4	0.040	3.55	14F/15F/15S
ss715587406	4	22,217,995	T/C	0.09	0.4	0.040	3.59	14F/15F/15S
ss715587428	4	23,132,742	T/C	0.09	0.45	0.048	4.16	14F/14S/15F/15S
ss715587431	4	23,251,720	T/C	0.08	0.45	0.047	4.13	14F/14S/15F/15S
ss715587436	4	23,567,358	T/C	0.10	0.44	0.049	4.22	14F/14S/15F/15S
ss715587463	4	24,276,060	G/T	0.09	0.4	0.040	3.55	14F/15F/15S
ss715587475	4	24,888,097	C/A	0.08	0.51	0.065	5.40	14F/14S/15F/15S
ss715587487	4	25,230,364	T/C	0.09	0.45	0.048	4.16	14F/14S/15F/15S
ss715587523	4	26,520,135	C/A	0.09	0.41	0.040	3.55	14F/15F/15S
ss715587524	4	26,558,203	A/G	0.09	0.44	0.048	4.16	14F/14S/15F/15S
ss715587534	4	26,886,535	T/C	0.08	0.45	0.047	4.13	14F/14S/15F/15S
ss715587541	4	27,207,244	A/G	0.08	0.37	0.033	3.04	14F/15F/15S
ss715587552	4	27,610,714	A/G	0.07	0.41	0.036	3.29	14F/15F/15S
ss715587553	4	27,662,543	T/C	0.07	0.41	0.036	3.29	14F/15F/15S
ss715587556	4	27,781,275	T/C	0.09	0.45	0.048	4.16	14F/14S/15F/15S
ss715587560	4	27,912,357	T/C	0.07	0.45	0.045	3.97	14F/14S/15F/15S
ss715587563	4	28,018,833	G/A	0.09	0.34	0.034	3.11	14F/14S/15S
ss715587634	4	30,543,000	T/C	0.10	0.37	0.034	3.12	14F/15F/15S
ss715587641	4	30,739,677	A/G	0.10	0.37	0.034	3.12	14F/15F/15S
ss715587653	4	31,038,638	T/C	0.10	0.37	0.034	3.12	14F/15F/15S
ss715587660	4	31,314,192	T/C	0.10	0.37	0.034	3.12	14F/15F/15S
ss715587684	4	32,065,239	T/C	0.10	0.37	0.034	3.12	14F/15F/15S
ss715587691	4	32,430,344	C/T	0.08	0.4	0.041	3.66	14F/15F/15S
ss715587713	4	32,933,642	G/A	0.08	0.44	0.048	4.21	14F/15F/15S
ss715587732	4	33,510,403	C/T	0.09	0.35	0.037	3.35	14F/14S/15S
ss715587735	4	33,590,548	G/A	0.09	0.41	0.043	3.84	14F/15F/15S
ss715609782	11	2,622,133	G/T	0.24	0.25	0.036	3.29	14F/14S/15S
ss715612171	12	32,474,069	G/A	0.23	0.18	0.034	3.10	14S/15F/15S
ss715630059	18	26,275,975	T/C	0.08	0.45	0.047	4.13	14F/14S/15F/15S
ss715630061	18	26,311,309	T/C	0.09	0.4	0.041	3.66	14F/14S/15F/15S
ss715630064	18	26,396,917	A/G	0.08	0.4	0.038	3.41	14F/15F/15S
ss715632339	18	60,741,380	A/G	0.48	-0.2	0.031	2.89	14S/15F/15S
ss715637058	20	24,384,755	G/A	0.09	0.34	0.037	3.34	14S/15F/15S
ss715637059	20	24,411,884	T/C	0.09	0.33	0.037	3.34	14S/15F/15S
ss715637063	20	24,611,709	A/G	0.09	0.34	0.037	3.34	14S/15F/15S
ss715637064	20	24,660,867	C/T	0.09	0.34	0.037	3.34	14S/15F/15S
ss715637077	20	25,096,146	T/C	0.09	0.33	0.037	3.34	14S/15F/15S
ss715637080	20	25,178,739	C/T	0.09	0.33	0.037	3.34	14S/15F/15S
ss715637081	20	25,207,060	G/A	0.09	0.33	0.037	3.34	14S/15F/15S
ss715637082	20	25,240,296	T/C	0.09	0.34	0.037	3.34	14S/15F/15S
ss715637086	20	25,487,504	G/A	0.09	0.34	0.037	3.34	14S/15F/15S
ss715637087	20	25,550,796	C/A	0.09	0.34	0.037	3.34	14S/15F/15S
ss715637088	20	25,585,353	A/G	0.09	0.33	0.037	3.34	14S/15F/15S
ss715637092	20	25,675,440	C/T	0.09	0.34	0.037	3.34	14S/15F/15S
ss715637101	20	25,951,602	T/G	0.09	0.34	0.037	3.34	14S/15F/15S
ss715637102	20	25,992,844	G/A	0.09	0.34	0.037	3.34	14S/15F/15S

Table 14 cont. The SNPs associated with seed height across all environments and in at least three of the four environments using GAPIT and Tassel software with a threshold of $-\log_{10}(P) \geq 2.5$.

SH SNP ID	Chr	Position	Allele	MAF[d]	Effect[c]	r^2	$-\log_{10}(p)$[b]	ENV[a]
ss715637104	20	26,067,212	C/A	0.09	0.31	0.036	3.27	14S/15F/15S
ss715637108	20	26,172,915	C/T	0.09	0.34	0.037	3.34	14S/15F/15S
ss715637109	20	26,206,802	C/T	0.09	0.34	0.037	3.34	14S/15F/15S
ss715637113	20	26,326,962	C/T	0.09	0.38	0.049	4.27	14F/14S/15F/15S
ss715637115	20	26,433,679	T/G	0.09	0.34	0.037	3.34	14S/15F/15S
ss715637117	20	26,500,747	A/G	0.09	0.34	0.037	3.34	14S/15F/15S
ss715637120	20	26,607,276	G/A	0.09	0.34	0.037	3.34	14S/15F/15S
ss715637125	20	26,759,063	A/C	0.09	0.33	0.037	3.34	14S/15F/15S
ss715637126	20	26,785,339	C/T	0.09	0.33	0.040	3.58	14S/15F/15S

[a] The individual environments the SNP was found to be associated with seed height
[b] Significance level of SNPs in the across all environments GWAS analysis
[c] Allele effect
[d] Minor allele frequency

Table 15. The SNPs associated with seed weight across all environments and in at least three of the four environments using GAPIT and Tassel software with a threshold of $-\log_{10}(p) \geq 2.5$.

SNP ID	Chr	Position	Allele	MAF[d]	Effect[c]	r^2	$-\log_{10}(p)$[b]	ENV[a]
ss715587326	4	19,903,602	A/G	0.09	3.35	0.037	3.35	14S/15F/15S
ss715587338	4	20,426,164	T/C	0.09	3.64	0.041	3.62	14S/15F/15S
ss715587357	4	20,796,034	A/G	0.09	3.13	0.033	3.07	14S/15F/15S
ss715587360	4	20,856,983	T/C	0.08	3.35	0.037	3.35	14S/15F/15S
ss715587370	4	21,252,334	A/G	0.09	3.40	0.037	3.33	14S/15F/15S
ss715587406	4	22,217,995	T/C	0.09	3.35	0.037	3.35	14S/15F/15S
ss715587428	4	23,132,742	T/C	0.09	3.58	0.041	3.63	14S/15F/15S
ss715587431	4	23,251,720	T/C	0.09	3.64	0.041	3.62	14S/15F/15S
ss715587436	4	23,567,358	T/C	0.09	3.53	0.042	3.72	14S/15F/15S
ss715587463	4	24,276,060	G/T	0.10	3.40	0.037	3.33	14S/15F/15S
ss715587475	4	24,888,097	C/A	0.08	4.39	0.063	5.29	14F/14S/15F/15S
ss715587487	4	25,230,364	T/C	0.09	3.58	0.041	3.63	14S/15F/15S
ss715587523	4	26,520,135	C/A	0.09	3.40	0.037	3.33	14S/15F/15S
ss715587524	4	26,558,203	A/G	0.09	3.58	0.041	3.63	14S/15F/15S
ss715587534	4	26,886,535	T/C	0.09	3.64	0.041	3.62	14S/15F/15S
ss715587552	4	27,610,714	A/G	0.07	3.92	0.044	3.86	14F/14S/15F/15S
ss715587553	4	27,662,543	T/C	0.07	3.92	0.044	3.86	14F/14S/15F/15S
ss715587556	4	27,781,275	T/C	0.09	3.58	0.041	3.63	14S/15F/15S
ss715587560	4	27,912,357	T/C	0.07	3.85	0.044	3.87	14S/15F/15S
ss715587691	4	32,430,344	C/T	0.09	3.30	0.037	3.38	14S/15F/15S
ss715587713	4	32,933,642	G/A	0.08	3.50	0.041	3.64	14S/15F/15S
ss715587735	4	33,590,548	G/A	0.09	3.26	0.037	3.36	14S/15F/15S
ss715609782	11	2,622,133	G/T	0.24	2.53	0.050	4.29	14F/14S/15S
ss715612567	12	35,454,411	A/G	0.45	1.64	0.031	2.90	14S/15F/15S
ss715630059	18	26,275,975	T/C	0.09	3.64	0.041	3.62	14S/15F/15S
ss715637058	20	24,384,755	G/A	0.09	2.93	0.037	3.38	14S/15F/15S
ss715637059	20	24,411,884	T/C	0.09	2.86	0.037	3.36	14S/15F/15S
ss715637063	20	24,611,709	A/G	0.09	2.93	0.037	3.38	14S/15F/15S
ss715637064	20	24,660,867	C/T	0.09	2.93	0.037	3.38	14S/15F/15S
ss715637077	20	25,096,146	T/C	0.09	2.86	0.037	3.35	14S/15F/15S
ss715637080	20	25,178,739	C/T	0.09	2.86	0.037	3.36	14S/15F/15S
ss715637082	20	25,240,296	T/C	0.09	2.93	0.037	3.38	14S/15F/15S
ss715637086	20	25,487,504	G/A	0.09	2.93	0.037	3.38	14S/15F/15S
ss715637087	20	25,550,796	C/A	0.09	2.93	0.037	3.38	14S/15F/15S
ss715637088	20	25,585,353	A/G	0.09	2.86	0.037	3.36	14S/15F/15S
ss715637092	20	25,675,440	C/T	0.09	2.93	0.037	3.38	14S/15F/15S
ss715637098	20	25,875,788	A/C	0.09	2.86	0.037	3.36	14S/15F/15S
ss715637101	20	25,951,602	T/G	0.09	2.93	0.037	3.38	14S/15F/15S
ss715637102	20	25,992,844	G/A	0.09	2.93	0.037	3.38	14S/15F/15S
ss715637104	20	26,067,212	C/A	0.09	2.85	0.040	3.61	14S/15F/15S
ss715637108	20	26,172,915	C/T	0.09	2.93	0.037	3.38	14S/15F/15S
ss715637109	20	26,206,802	C/T	0.09	2.93	0.037	3.38	14S/15F/15S
ss715637113	20	26,326,962	C/T	0.09	3.33	0.050	4.35	14F/14S/15F/15S
ss715637115	20	26,433,679	T/G	0.09	2.93	0.037	3.38	14S/15F/15S
ss715637117	20	26,500,747	A/G	0.09	2.93	0.037	3.38	14S/15F/15S
ss715637120	20	26,607,276	G/A	0.09	2.93	0.037	3.38	14S/15F/15S
ss715637125	20	26,759,063	A/C	0.09	2.86	0.037	3.36	14S/15F/15S

[a] The individual environments the SNP was found to be associated with seed weight
[b] Significance level of SNPs in the across all environments GWAS analysis
[c] Allele effect
[d] Minor allele frequency

Table 16. The SNPs associated with seed volume across all environments and in at least three of the four environments using GAPIT and Tassel software with a threshold of $-\log_{10}(p) \geq 2.5$.

VOL SNP ID	Chr.	Position	Allele	MAF[d]	Effect[c]	r^2	$-\log_{10}(p)$[b]	ENV[a]
ss715587338	4	20,426,164	T/C	0.085	51.20	0.036	3.26	14S/15F/15S
ss715587428	4	23,132,742	T/C	0.087	50.60	0.036	3.28	14S/15F/15S
ss715587431	4	23,251,720	T/C	0.085	51.20	0.036	3.26	14S/15F/15S
ss715587436	4	23,567,358	T/C	0.096	50.20	0.038	3.39	14S/15F/15S
ss715587475	4	24,888,097	C/A	0.082	59.00	0.049	4.25	14F/14S/15F/15S
ss715587487	4	25,230,364	T/C	0.087	50.60	0.036	3.28	14S/15F/15S
ss715587524	4	26,558,203	A/G	0.087	50.55	0.036	3.28	14S/15F/15S
ss715587534	4	26,886,535	T/C	0.085	51.10	0.036	3.26	14S/15F/15S
ss715587556	4	27,781,275	T/C	0.087	50.60	0.036	3.28	14S/15F/15S
ss715587560	4	27,912,357	T/C	0.070	54.50	0.039	3.50	14S/15F/15S
ss715587713	4	32,933,642	G/A	0.082	46.60	0.031	2.89	14F/15F/15S
ss715609782	11	2,622,133	G/T	0.245	39.61	0.053	4.51	14F/14S/15S
ss715609785	11	2,626,602	A/G	0.257	32.46	0.034	3.09	14F/14S/15S
ss715630059	18	26,275,975	T/C	0.085	51.20	0.036	3.26	14S/15F/15S
ss715637058	20	24,384,755	G/A	0.090	44.90	0.039	3.46	14S/15F/15S
ss715637059	20	24,411,884	T/C	0.093	43.80	0.038	3.44	14S/15F/15S
ss715637063	20	24,611,709	A/G	0.090	44.89	0.039	3.46	14S/15F/15S
ss715637064	20	24,660,867	C/T	0.090	44.89	0.039	3.46	14S/15F/15S
ss715637077	20	25,096,146	T/C	0.095	43.70	0.038	3.43	14S/15F/15S
ss715637080	20	25,178,739	C/T	0.093	43.77	0.038	3.44	14S/15F/15S
ss715637081	20	25,207,060	G/A	0.093	43.80	0.038	3.44	14S/15F/15S
ss715637082	20	25,240,296	T/C	0.090	44.90	0.039	3.46	14S/15F/15S
ss715637086	20	25,487,504	G/A	0.090	44.90	0.039	3.46	14S/15F/15S
ss715637087	20	25,550,796	C/A	0.090	44.90	0.039	3.46	14S/15F/15S
ss715637088	20	25,585,353	A/G	0.093	43.77	0.038	3.44	14S/15F/15S
ss715637092	20	25,675,440	C/T	0.090	44.89	0.039	3.46	14S/15F/15S
ss715637098	20	25,875,788	A/C	0.093	43.77	0.038	3.44	14S/15F/15S
ss715637101	20	25,951,602	T/G	0.090	44.90	0.039	3.46	14S/15F/15S
ss715637102	20	25,992,844	G/A	0.090	44.90	0.039	3.46	14S/15F/15S
ss715637104	20	26,067,212	C/A	0.093	42.10	0.039	3.49	14S/15F/15S
ss715637108	20	26,172,915	C/T	0.090	44.89	0.039	3.46	14S/15F/15S
ss715637109	20	26,206,802	C/T	0.090	44.89	0.039	3.46	14S/15F/15S
ss715637113	20	26,326,962	C/T	0.090	47.57	0.044	3.87	14S/15F/15S
ss715637115	20	26,433,679	T/G	0.090	44.90	0.039	3.46	14S/15F/15S
ss715637117	20	26,500,747	A/G	0.090	44.89	0.039	3.46	14S/15F/15S
ss715637120	20	26,607,276	G/A	0.090	44.90	0.039	3.46	14S/15F/15S
ss715637125	20	26,759,063	A/C	0.093	43.77	0.038	3.44	14S/15F/15S
ss715637126	20	26,785,339	C/T	0.093	43.94	0.039	3.51	14S/15F/15S

[a] The individual environments the SNP was found to be associated with seed volume
[b] Significance level of SNPs in the across all environments GWAS analysis
[c] Allele effect
[d] Minor allele frequency

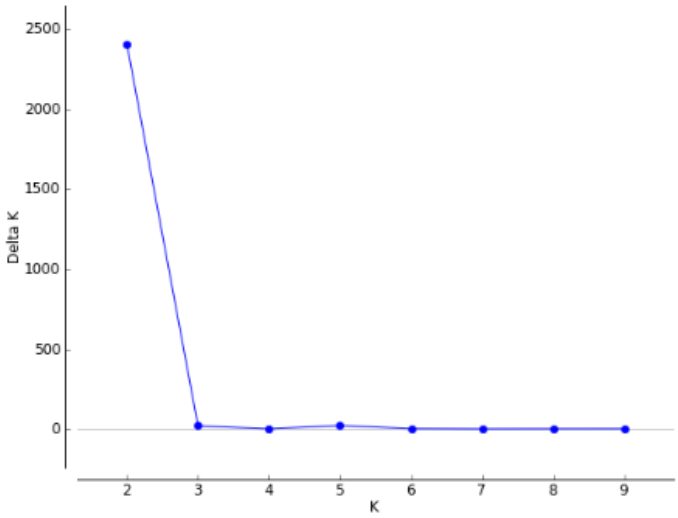
Table 17. The number of favorable alleles associated with seed weight found in 60 entries using a 6K SNP chip.

Favorable Alleles	Number of Accessions	AVG SW[a] (g/100s)
8	15	25.24
7	8	23.35
6	8	21.57
3	2	24.55
2	8	22.32
1	13	21.11
0	6	17.62

a Average seed weight
b The number of favorable alleles associated with seed weight out of 47 SNPs associated with seed weight found in the 50K SNP chip

Figure 1. Peak Delta K from STRUCTURE Harvester.

A. Delta K of the 343 GRIN accessions.



B. Delta K of the 31 breeding lines and 29 accessions.

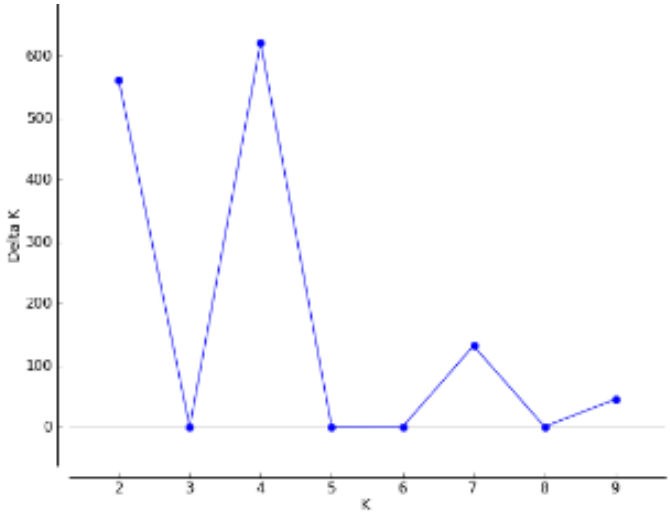


Figure 2. Phylogenetic tree of the 343 GRIN accessions.

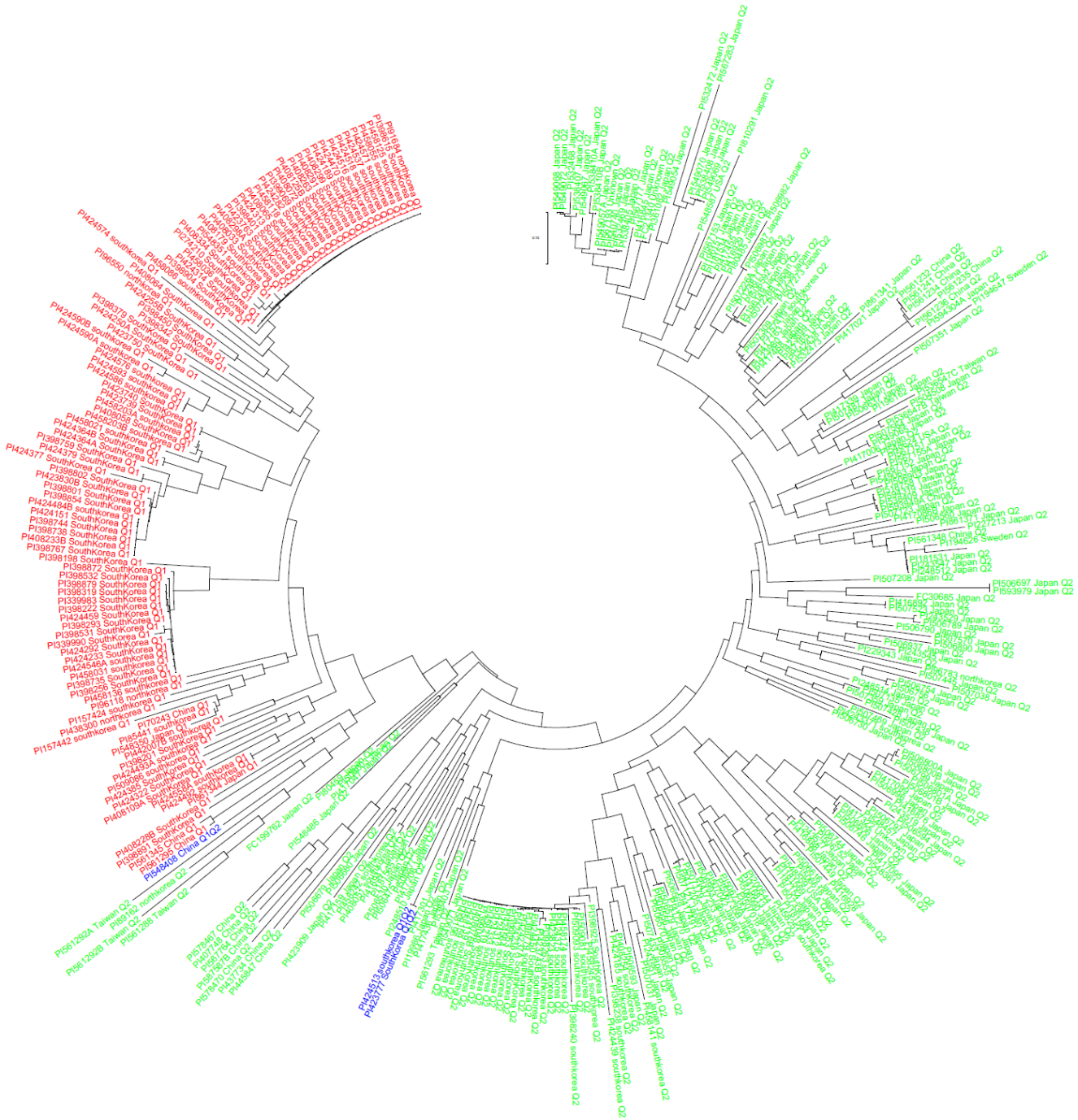


Figure 3. Phylogenetic tree of 31 breeding lines and 29 GRIN accessions.

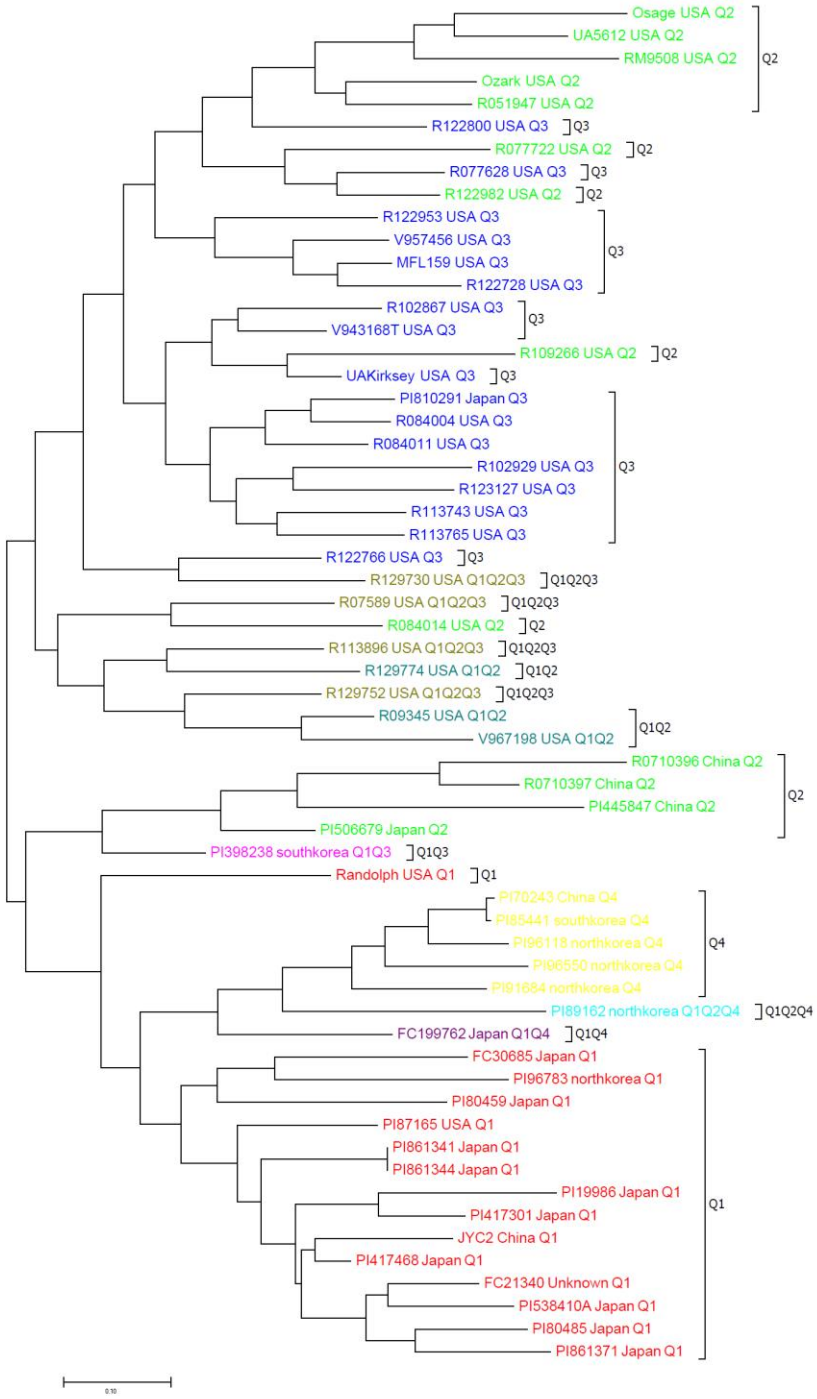


Figure 4. Linkage disequilibrium plot of the 343 GRIN accessions.

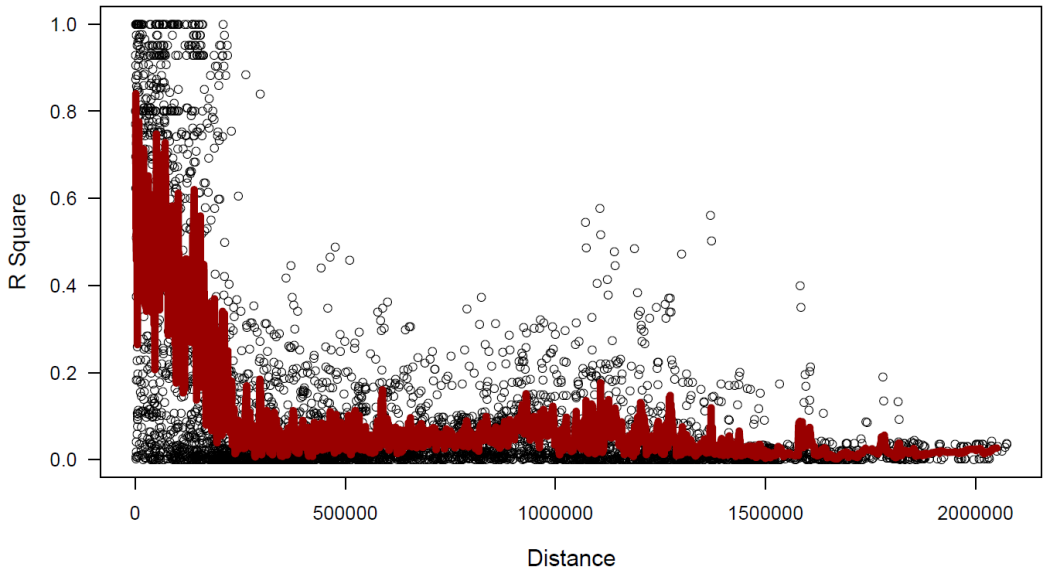
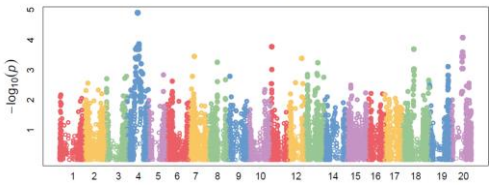
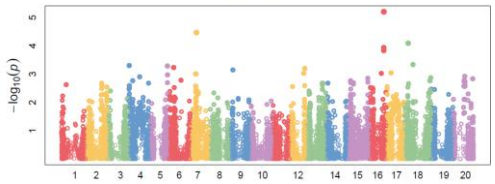


Figure 5. Manhattan plots displaying SNPs associated with seed weight and size trait across all environments (AAE) with a significance level of $-\log_{10}(p) > 2.5$.

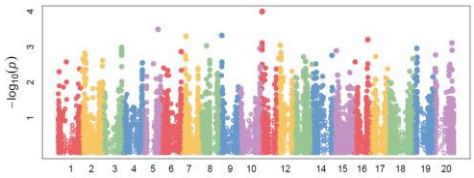
A. Seed Weight



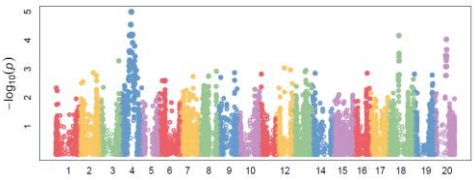
B. Seed Length



C. Seed Breadth



D. Seed Height



E. Volume

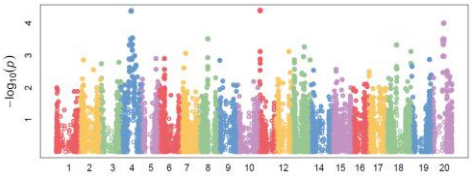


Figure 6. SNPs associated with seed weight and size traits ($-\log_{10}(p) > 2.5$) across all environments and in at least three out of four environments.

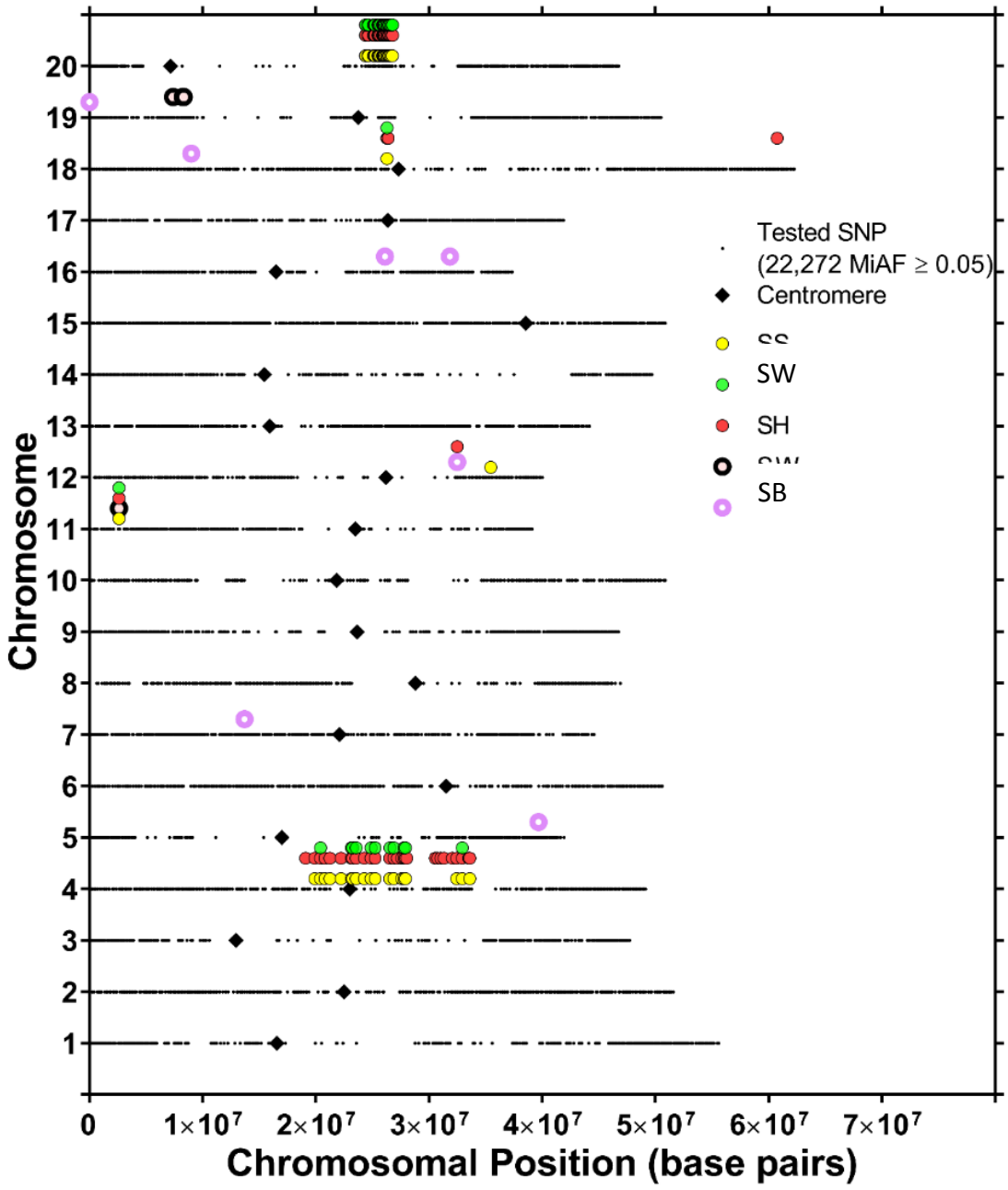
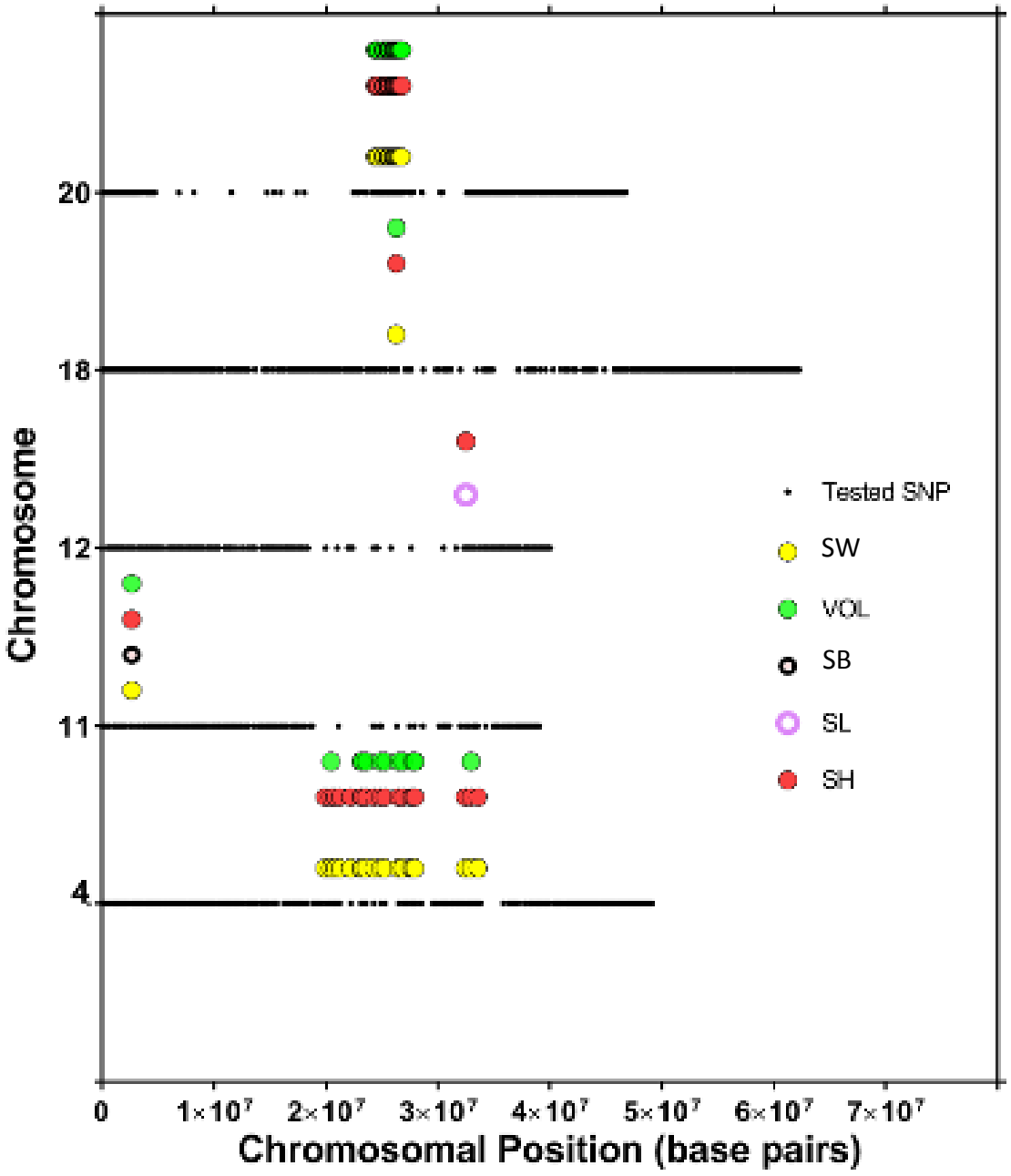


Figure 7. SNPs associated with more than one trait ($-\log_{10}(p) > 2.5$) located on chromosomes 4, 11, 12, 18, and 20.



Supplementary Table 1. Accession name, maturity group (MG), country of origin, and phenotypic data, for all 343 soybean accessions acquired from the GRIN database. Traits are: seed weight (SW), Seed Length (SL), seed breadth (SB), seed height (SH), and volume.

Accession	MG	Country	SW (g/100s)	SL (mm)	SB (mm)	SH (mm)	Volume (mm^3)
PI 561232	1	China	24.21	8.09	6.30	7.52	383.82
PI 561235	1	China	21.42	7.83	5.97	7.19	336.41
PI 561241	1	China	22.91	7.74	6.35	7.38	362.76
PI 561295	1	China	23.73	9.05	5.94	7.41	397.19
PI 561345	1	China	25.37	8.90	6.34	7.51	423.72
PI 561348	1	China	24.30	8.95	5.37	8.06	388.21
PI 593949A	1	China	26.41	7.93	6.77	7.92	426.65
PI 561234	2	China	22.86	7.97	6.35	7.46	376.69
PI 561236	2	China	23.24	7.79	6.32	7.51	369.36
PI 548408	4	China	24.07	8.69	5.84	7.47	379.23
PI 70243	4	China	28.41	8.64	6.56	7.74	439.57
PI 407748	5	China	24.97	8.71	6.04	7.37	387.97
PI 437734	5	China	23.92	8.50	5.84	7.24	360.53
PI 567764	5	China	24.64	8.31	6.14	7.77	396.20
PI 578467	6	China	27.79	8.90	6.18	7.89	436.05
PI 578470	7	China	26.00	10.87	4.44	7.90	383.95
PI 587587B	8	China	26.91	9.37	5.66	7.55	402.05
PI 445847	9	China	19.21	9.47	4.82	6.31	288.98
PI 506697	000	Japan	13.60	7.16	4.80	6.19	212.91
PI 593979	000	Japan	13.20	7.18	4.74	6.18	210.75
FC 30685	00	Japan	21.88	8.47	5.40	7.38	338.32
PI 538408	00	Japan	19.30	7.14	6.00	7.15	310.02
PI 549054	00	Japan	20.82	8.35	6.07	7.37	375.64
PI 567283	00	Japan	21.83	8.03	6.17	7.58	376.68
PI 181531	0	Japan	22.31	8.07	5.94	7.32	351.19
PI 243547	0	Japan	21.68	7.92	5.88	7.28	339.54
PI 248512	0	Japan	21.69	7.96	5.94	7.24	342.61
PI 361082	0	Japan	23.49	8.27	6.23	7.52	389.15
PI 416845	0	Japan	25.13	8.50	6.15	7.78	407.28
PI 417095	0	Japan	22.15	8.62	5.24	7.42	337.11
PI 507038	0	Japan	18.78	7.69	5.56	6.97	301.40
PI 507351	0	Japan	22.23	8.32	5.50	7.38	338.83
PI 549069	0	Japan	17.90	7.91	5.48	6.78	294.49
PI 549070	0	Japan	24.19	8.01	6.45	7.50	387.91
PI 567177	0	Japan	20.83	8.23	5.82	7.22	345.37
PI 532468	1	Japan	25.99	8.78	6.39	7.53	422.54
PI 532469	1	Japan	17.29	8.07	5.37	6.83	299.78
PI 538403	1	Japan	24.69	8.06	6.26	7.64	385.51
PI 538405	1	Japan	21.31	8.26	5.75	7.17	340.61
PI 538407	1	Japan	26.07	8.61	6.32	7.67	417.44
PI 538409	1	Japan	26.77	7.95	6.63	7.88	416.60
PI 538410A	1	Japan	23.26	8.60	6.17	7.38	391.17
PI 538410B	1	Japan	24.05	8.79	6.06	7.40	393.58
PI 540740	1	Japan	22.03	8.10	6.08	7.34	362.33
PI 549057A	1	Japan	23.69	8.49	6.20	7.49	394.42
PI 549067	1	Japan	23.96	8.70	6.00	7.31	381.65
PI 549068	1	Japan	25.21	8.70	6.33	7.58	417.57
PI 549072	1	Japan	24.80	8.70	6.39	7.57	423.12
PI 567155A	1	Japan	23.80	8.36	5.83	7.54	367.80
PI 594304A	1	Japan	21.85	8.04	5.92	7.28	347.16
PI 594319	1	Japan	25.28	8.04	6.63	7.82	417.49

Supplementary Table 1 (cont.). Accession name, maturity group (MG), country of origin, and phenotypic data, for all 343 soybean accessions acquired from the GRIN database. Traits are: seed weight (SW), Seed Length (SL), seed breadth (SB), seed height (SH), and volume.

Accession	MG	Country	SW (g/100s)	SL (mm)	SB (mm)	SH (mm)	Volume (mm^3)
PI 181534	2	Japan	26.35	9.13	5.60	8.27	421.67
PI 196151	2	Japan	31.00	8.87	6.86	8.07	492.48
PI 227213	2	Japan	19.35	8.01	5.70	7.34	335.68
PI 416929	2	Japan	22.93	8.13	5.93	7.57	369.11
PI 417436	2	Japan	19.49	7.58	5.54	7.00	294.72
PI 417455	2	Japan	25.66	9.05	5.52	8.14	405.01
PI 532472	2	Japan	27.22	8.26	6.75	8.04	449.74
PI 532473	2	Japan	32.38	8.93	7.20	8.67	559.52
PI 549071	2	Japan	23.47	8.49	5.92	7.46	374.03
PI 567151	2	Japan	23.33	8.44	5.76	7.53	366.50
PI 567152	2	Japan	24.42	8.17	6.15	7.64	384.12
PI 567153	2	Japan	27.70	9.21	5.69	8.09	425.28
PI 80485	2	Japan	26.79	9.55	5.44	8.22	425.98
PI 861371	2	Japan	24.10	8.66	5.71	7.72	381.43
PI 196149	3	Japan	31.44	9.19	6.84	8.47	532.78
PI 196162	3	Japan	25.19	8.59	6.17	7.75	410.63
PI 342438	3	Japan	26.41	8.35	6.39	7.70	411.29
PI 416892	3	Japan	31.24	9.74	6.03	8.06	473.53
PI 423899	3	Japan	29.06	8.73	6.67	8.02	467.43
PI 504508	3	Japan	20.48	7.78	5.80	7.14	324.17
PI 506592	3	Japan	33.41	9.00	6.84	8.47	521.54
PI 506637	3	Japan	31.07	8.75	6.46	8.22	465.05
PI 506790	3	Japan	29.93	8.67	6.67	7.93	458.95
PI 506799	3	Japan	28.10	11.69	4.15	8.16	395.80
PI 506800A	3	Japan	29.79	12.53	4.05	8.36	424.25
PI 506800B	3	Japan	29.18	12.40	4.06	8.22	413.17
PI 506801A	3	Japan	26.21	10.91	4.45	8.20	398.14
PI 506801B	3	Japan	28.04	11.06	4.72	8.07	421.02
PI 506982	3	Japan	26.00	9.35	5.52	7.91	407.35
PI 506987	3	Japan	26.58	8.89	5.93	8.10	429.68
PI 507226A	3	Japan	29.25	9.72	5.76	8.17	451.20
PI 507226B	3	Japan	31.09	8.85	6.95	8.51	526.18
PI 507273	3	Japan	23.08	8.49	5.91	7.79	390.69
PI 507487	3	Japan	28.70	8.59	6.59	8.11	459.18
PI 507523	3	Japan	31.38	9.81	5.88	8.04	463.31
PI 507570	3	Japan	30.59	8.84	6.78	8.18	490.65
PI 548361	3	Japan	28.68	9.86	5.29	8.42	441.07
PI 80459	3	Japan	25.06	9.07	5.75	7.54	394.26
FC 199762	4	Japan	19.87	8.41	5.25	6.90	304.44
PI 124871	4	Japan	32.41	9.31	6.39	8.40	499.65
PI 19986	4	Japan	25.80	8.14	6.35	7.59	392.75
PI 229343	4	Japan	31.50	8.81	6.70	8.13	480.42
PI 243519	4	Japan	34.45	11.35	5.31	8.76	529.31
PI 243527	4	Japan	30.67	10.96	4.90	8.32	447.41
PI 243529	4	Japan	29.03	8.39	6.81	8.10	463.57
PI 243545	4	Japan	26.33	9.29	5.62	7.61	396.25
PI 243551	4	Japan	31.53	8.80	6.93	8.19	499.25
PI 248514	4	Japan	29.35	8.80	6.39	8.39	473.39
PI 416888	4	Japan	29.79	10.79	4.88	8.27	435.33
PI 417006	4	Japan	30.15	9.00	6.21	7.96	445.08
PI 417021	4	Japan	23.12	8.27	5.94	7.28	357.38

Supplementary Table 1 (cont.). Accession name, maturity group (MG), country of origin, and phenotypic data, for all 343 soybean accessions acquired from the GRIN database. Traits are: seed weight (SW), Seed Length (SL), seed breadth (SB), seed height (SH), and volume.

Accession	MG	Country	SW (g/100s)	SL (mm)	SB (mm)	SH (mm)	Volume (mm^3)
PI 417086B	4	Japan	26.44	8.79	6.21	7.74	423.10
PI 417163	4	Japan	31.48	8.83	6.69	8.25	487.28
PI 417233	4	Japan	29.65	8.27	7.08	8.05	471.42
PI 417238	4	Japan	23.98	8.24	6.54	7.41	399.69
PI 417301	4	Japan	28.58	8.20	6.78	7.89	439.05
PI 417339	4	Japan	26.27	8.36	6.21	7.66	398.09
PI 417468	4	Japan	29.39	8.61	6.58	8.12	460.14
PI 423980	4	Japan	28.69	8.47	6.48	7.98	439.46
PI 506560	4	Japan	27.77	8.48	6.38	7.82	423.22
PI 506789	4	Japan	28.77	8.42	6.91	7.94	461.42
PI 506903	4	Japan	32.59	8.28	7.21	8.31	496.75
PI 506937	4	Japan	28.66	8.50	6.73	7.74	443.05
PI 506993	4	Japan	30.04	8.79	6.71	7.98	470.67
PI 507123	4	Japan	26.77	8.62	6.25	7.81	420.71
PI 507179	4	Japan	31.84	8.40	7.04	8.20	484.79
PI 507309	4	Japan	29.33	8.46	6.69	8.06	456.16
PI 507445	4	Japan	29.34	8.89	6.40	7.90	450.44
PI 507449	4	Japan	26.75	10.53	4.61	7.97	386.47
PI 507564	4	Japan	33.16	9.25	6.69	7.89	485.94
PI 548350	4	Japan	29.57	8.76	6.92	7.81	473.49
PI 549063	4	Japan	26.55	8.66	5.77	7.94	396.76
PI 549065	4	Japan	33.89	9.35	6.63	8.47	525.63
PI 810291	4	Japan	21.56	7.96	5.81	7.49	347.96
PI 861341	4	Japan	24.22	8.55	5.94	7.31	370.97
PI 861344	4	Japan	25.84	9.11	5.72	7.24	378.11
PI 417159	5	Japan	20.19	7.70	5.73	7.04	311.28
PI 417205	5	Japan	30.16	8.52	6.89	8.17	480.98
PI 417491	5	Japan	27.20	9.89	5.32	8.05	423.09
PI 506594	5	Japan	34.19	11.02	5.32	8.83	518.12
PI 506730	5	Japan	26.30	8.32	6.34	7.72	409.97
PI 506746	5	Japan	37.57	11.99	5.17	9.28	574.04
PI 506752	5	Japan	34.91	11.49	4.95	8.89	505.53
PI 506797	5	Japan	19.29	8.79	4.91	7.04	304.31
PI 506890	5	Japan	28.05	8.25	6.77	7.64	428.37
PI 507031	5	Japan	30.70	8.46	6.83	8.08	467.57
PI 507121	5	Japan	26.13	8.25	6.59	8.08	442.11
PI 507135	5	Japan	28.85	9.87	6.46	7.94	507.56
PI 507433	5	Japan	31.17	9.07	6.51	8.12	479.77
PI 416876	6	Japan	36.48	13.49	3.97	9.63	515.83
PI 417099	6	Japan	35.46	12.73	4.37	8.93	496.99
PI 506530	6	Japan	26.04	8.54	6.10	7.73	404.08
PI 506569	6	Japan	32.07	9.37	6.46	8.28	502.49
PI 506593	6	Japan	25.16	8.20	6.33	7.63	398.84
PI 506606	6	Japan	34.78	11.59	5.01	9.15	531.48
PI 506744	6	Japan	34.03	11.32	4.91	9.27	515.55
PI 506754	6	Japan	20.54	7.48	5.86	7.12	311.95
PI 506996	6	Japan	20.40	9.45	4.59	6.86	297.29
PI 507208	6	Japan	26.75	8.17	6.68	7.72	421.83
PI 507428	6	Japan	30.39	9.05	6.53	8.16	483.64
PI 507438	6	Japan	28.72	8.77	6.48	7.97	454.39
PI 507469	6	Japan	20.84	8.11	5.76	7.15	334.04

Supplementary Table 1 (cont.). Accession name, maturity group (MG), country of origin, and phenotypic data, for all 343 soybean accessions acquired from the GRIN database. Traits are: seed weight (SW), Seed Length (SL), seed breadth (SB), seed height (SH), and volume.

Accession	MG	Country	SW (g/100s)	SL (mm)	SB (mm)	SH (mm)	Volume (mm^3)
PI 548457	6	Japan	31.75	10.39	5.25	8.47	461.74
PI 548486	6	Japan	28.15	9.39	5.71	7.89	424.39
PI 181565	7	Japan	27.89	8.62	6.28	8.04	437.86
PI 181569	7	Japan	26.82	9.62	5.41	7.92	412.25
PI 187154	7	Japan	30.30	8.94	6.41	8.08	465.40
PI 200544	7	Japan	35.87	9.59	7.09	8.67	590.56
PI 248510	7	Japan	33.71	10.60	5.55	8.35	491.57
PI 416813	7	Japan	33.31	9.25	6.49	8.45	508.82
PI 416928	7	Japan	28.78	10.78	4.55	8.39	411.94
PI 416947	7	Japan	32.86	9.35	6.42	8.39	506.63
PI 417047	7	Japan	32.62	10.55	5.55	8.44	494.15
PI 417206	7	Japan	31.87	9.57	6.47	8.18	506.67
PI 417270	7	Japan	33.76	9.08	6.88	8.61	539.19
PI 423909	7	Japan	17.79	8.27	5.14	6.56	279.18
PI 506475	7	Japan	29.29	10.55	4.94	8.50	437.74
PI 506555	7	Japan	30.77	9.10	6.44	8.20	482.49
PI 506556	7	Japan	35.95	9.11	6.84	8.61	538.17
PI 506570	7	Japan	33.22	8.95	6.61	8.50	503.08
PI 506603	7	Japan	31.85	10.85	4.98	8.81	476.43
PI 506616	7	Japan	29.11	10.33	4.84	8.37	419.64
PI 506618	7	Japan	27.00	10.17	4.76	8.38	406.90
PI 506735A	7	Japan	31.48	9.17	6.35	8.32	483.96
PI 506735B	7	Japan	28.39	9.00	6.30	7.65	433.42
PI 506756	7	Japan	20.20	7.66	5.91	7.28	330.37
PI 506877	7	Japan	33.76	8.94	6.62	8.21	488.45
PI 506990	7	Japan	39.22	10.18	6.62	8.74	589.22
PI 507042	7	Japan	30.56	9.43	6.31	8.13	486.03
PI 507336	7	Japan	34.58	9.52	6.80	8.54	554.23
PI 507359	7	Japan	31.10	9.60	6.42	8.01	495.70
PI 181564	8	Japan	36.46	10.38	6.52	8.41	569.26
PI 506579	8	Japan	32.22	9.18	6.84	8.66	547.58
PI 506679	8	Japan	27.99	9.41	6.03	7.68	437.08
PI 506680	8	Japan	31.04	9.94	6.41	7.65	489.63
PI 507018	8	Japan	26.44	8.20	6.52	8.11	431.82
PI 89162	3	North Korea	23.55	9.25	5.44	7.30	367.65
PI 438300	4	North Korea	29.60	8.88	6.60	7.84	459.81
PI 91684	4	North Korea	26.36	8.38	6.35	7.44	397.18
PI 96118	4	North Korea	26.04	8.31	6.37	7.61	402.98
PI 96550	4	North Korea	26.39	8.90	6.21	7.53	416.31
PI 96783	4	North Korea	19.84	7.99	5.85	6.78	317.94
PI 157419	4	South Korea	29.00	9.11	6.41	8.32	485.78
PI 157424	4	South Korea	28.27	8.70	6.41	7.80	435.54
PI 157442	4	South Korea	23.03	8.01	6.26	7.35	368.68
PI 274210	4	South Korea	28.64	8.66	6.60	7.84	449.93
PI 339983	4	South Korea	27.42	8.42	6.24	8.06	424.08
PI 339990	4	South Korea	27.60	8.66	6.16	8.10	432.25
PI 398198	4	South Korea	26.88	9.15	5.76	8.00	421.31
PI 398201	4	South Korea	26.79	8.40	6.16	8.02	416.39
PI 398222	4	South Korea	27.82	8.46	6.24	8.19	433.35
PI 398256	4	South Korea	27.24	8.55	6.23	8.09	431.30
PI 398293	4	South Korea	29.15	8.57	6.49	8.29	461.53

Supplementary Table 1 (cont.). Accession name, maturity group (MG), country of origin, and phenotypic data, for all 343 soybean accessions acquired from the GRIN database. Traits are: seed weight (SW), Seed Length (SL), seed breadth (SB), seed height (SH), and volume.

Accession	MG	Country	SW (g/100s)	SL (mm)	SB (mm)	SH (mm)	Volume (mm^3)
PI 398319	4	South Korea	28.66	8.61	6.29	8.10	438.71
PI 398342	4	South Korea	27.96	8.45	6.46	8.07	440.19
PI 398379	4	South Korea	26.48	9.34	5.79	7.79	421.87
PI 398401	4	South Korea	30.76	8.68	6.78	7.97	469.74
PI 398450	4	South Korea	28.59	8.89	6.34	8.12	457.49
PI 398531	4	South Korea	26.79	8.39	6.19	8.11	421.67
PI 398532	4	South Korea	27.71	8.48	6.38	8.12	439.43
PI 398615	4	South Korea	29.37	8.69	6.68	7.98	464.05
PI 398735	4	South Korea	26.36	8.45	6.07	8.04	411.98
PI 398738	4	South Korea	30.94	9.31	6.20	8.45	488.20
PI 398744	4	South Korea	28.91	9.14	6.01	8.35	458.60
PI 398759	4	South Korea	27.65	9.06	6.00	8.30	451.42
PI 398767	4	South Korea	30.12	9.14	6.20	8.43	478.87
PI 398801	4	South Korea	30.16	9.14	6.17	8.40	474.59
PI 398802	4	South Korea	24.56	8.63	5.67	7.74	379.14
PI 398854	4	South Korea	29.30	9.02	6.09	8.25	454.12
PI 398872	4	South Korea	26.61	8.43	6.18	8.02	417.86
PI 398879	4	South Korea	28.06	8.65	6.26	8.18	442.44
PI 398891	4	South Korea	28.43	8.35	6.71	7.93	445.39
PI 398904	4	South Korea	29.27	8.55	6.69	7.81	447.56
PI 398925	4	South Korea	22.39	7.78	6.05	7.16	337.79
PI 399048	4	South Korea	30.23	9.28	6.17	8.07	462.61
PI 399053	4	South Korea	27.35	9.14	5.95	7.86	427.75
PI 399069	4	South Korea	28.22	8.32	6.63	7.84	432.91
PI 408033	4	South Korea	29.19	8.41	6.83	7.86	452.68
PI 408058	4	South Korea	31.94	9.30	6.42	8.39	501.18
PI 408064	4	South Korea	26.86	8.88	5.96	7.76	410.96
PI 408065	4	South Korea	29.47	9.12	6.05	8.46	466.96
PI 408072	4	South Korea	28.86	8.51	6.64	7.92	449.35
PI 408076B	4	South Korea	31.78	9.21	6.46	8.12	483.37
PI 408109A	4	South Korea	26.72	8.55	6.26	7.84	419.81
PI 408125B	4	South Korea	29.77	8.59	6.79	7.99	466.62
PI 408179	4	South Korea	28.44	8.44	6.72	7.93	450.40
PI 408228B	4	South Korea	16.86	7.65	5.22	6.49	259.44
PI 408233B	4	South Korea	29.93	9.28	6.04	8.33	467.66
PI 408263	4	South Korea	28.93	8.61	6.68	7.95	457.65
PI 408291	4	South Korea	31.03	8.75	6.88	8.06	486.22
PI 408298A	4	South Korea	29.42	8.67	6.70	8.02	466.16
PI 408299	4	South Korea	27.62	8.33	6.68	7.78	433.57
PI 408334	4	South Korea	31.40	8.71	6.81	8.07	480.08
PI 423739	4	South Korea	24.76	8.73	5.83	7.61	387.86
PI 423740	4	South Korea	26.15	8.72	5.88	7.69	394.31
PI 423750	4	South Korea	26.61	9.45	5.53	7.85	409.69
PI 423763	4	South Korea	28.39	8.46	6.79	7.75	445.04
PI 423777	4	South Korea	26.47	8.42	6.44	7.59	411.92
PI 423830B	4	South Korea	27.38	9.01	5.96	8.27	445.16
PI 424151	4	South Korea	29.87	9.27	6.08	8.29	468.85
PI 424189	4	South Korea	30.09	8.68	6.67	7.90	459.32
PI 424233	4	South Korea	27.95	8.65	6.24	8.16	441.84
PI 424237A	4	South Korea	29.65	9.12	6.48	8.02	474.22
PI 424238	4	South Korea	32.93	9.24	6.60	8.14	495.48

Supplementary Table 1 (cont.). Accession name, maturity group (MG), country of origin, and phenotypic data, for all 343 soybean accessions acquired from the GRIN database. Traits are: seed weight (SW), Seed Length (SL), seed breadth (SB), seed height (SH), and volume.

Accession	MG	Country	SW (g/100s)	SL (mm)	SB (mm)	SH (mm)	Volume (mm^3)
PI 424250A	4	South Korea	27.19	9.14	5.89	7.89	426.62
PI 424255B	4	South Korea	24.51	8.35	6.20	7.39	382.93
PI 424282	4	South Korea	29.01	8.60	6.72	7.94	461.52
PI 424292	4	South Korea	25.45	8.35	6.06	7.90	400.23
PI 424313	4	South Korea	29.65	8.63	6.68	8.03	464.49
PI 424314	4	South Korea	29.19	8.54	6.69	7.90	452.69
PI 424322	4	South Korea	27.22	9.09	5.93	8.02	432.67
PI 424364A	4	South Korea	30.35	9.27	5.84	8.44	458.33
PI 424364B	4	South Korea	29.61	9.17	5.84	8.29	444.14
PI 424377	4	South Korea	28.01	8.91	5.88	8.22	430.84
PI 424379	4	South Korea	30.07	9.09	6.02	8.34	457.85
PI 424385	4	South Korea	27.10	8.64	6.27	7.76	421.67
PI 424450	4	South Korea	32.32	9.35	6.42	8.13	489.56
PI 424459	4	South Korea	29.24	8.64	6.36	8.26	453.52
PI 424470	4	South Korea	30.85	8.68	6.83	8.04	478.05
PI 424484B	4	South Korea	28.98	9.05	6.04	8.18	447.09
PI 424492	4	South Korea	27.79	8.96	6.14	7.95	437.71
PI 424493A	4	South Korea	27.53	9.25	5.71	8.15	430.19
PI 424513	4	South Korea	25.28	8.30	6.28	7.47	389.83
PI 424516	4	South Korea	29.14	8.58	6.68	7.76	446.55
PI 424518	4	South Korea	30.15	8.53	6.73	7.95	457.39
PI 424537	4	South Korea	30.34	8.52	6.70	7.93	454.41
PI 424546A	4	South Korea	27.45	8.67	6.32	8.24	451.90
PI 424558A	4	South Korea	27.06	9.03	5.86	7.89	418.18
PI 424564	4	South Korea	20.96	7.95	5.70	6.98	317.37
PI 424571	4	South Korea	29.43	8.50	6.71	7.85	449.80
PI 424574	4	South Korea	17.51	7.15	5.67	6.67	271.19
PI 424590A	4	South Korea	27.43	9.36	5.97	8.00	447.29
PI 424590B	4	South Korea	28.80	9.53	5.81	7.95	441.00
PI 442007B	4	South Korea	27.77	8.56	6.33	7.86	426.36
PI 458021	4	South Korea	33.71	9.33	6.71	8.22	515.08
PI 458031	4	South Korea	27.06	8.48	6.21	8.12	428.67
PI 458036	4	South Korea	28.27	8.50	6.71	7.96	454.42
PI 458055	4	South Korea	31.42	8.79	6.94	8.12	497.37
PI 458086	4	South Korea	26.88	8.43	6.32	7.75	414.07
PI 458118	4	South Korea	31.32	8.59	6.99	8.01	483.05
PI 458125	4	South Korea	28.22	8.51	6.85	7.94	465.83
PI 458136	4	South Korea	27.52	8.55	6.41	7.77	426.36
PI 458141	4	South Korea	32.00	8.55	7.12	8.22	502.09
PI 458203A	4	South Korea	33.48	9.38	6.60	8.40	520.23
PI 458203B	4	South Korea	33.05	9.36	6.57	8.46	519.85
PI 548351	4	South Korea	30.56	8.59	6.75	8.03	466.13
PI 85441	4	South Korea	28.53	8.67	6.57	7.70	439.07
PI 398238	5	South Korea	25.94	9.09	6.25	7.83	446.02
PI 398240	5	South Korea	23.91	8.67	5.83	7.36	372.95
PI 398263	5	South Korea	25.77	8.54	6.07	7.52	391.72
PI 398779	5	South Korea	31.72	9.19	6.54	8.12	489.49
PI 423743B	5	South Korea	29.12	8.89	6.27	8.03	449.20
PI 423823	5	South Korea	20.65	7.67	5.99	6.94	319.63
PI 424312	5	South Korea	25.27	8.91	6.28	7.43	414.73

Supplementary Table 1 (cont.). Accession name, maturity group (MG), country of origin, and phenotypic data, for all 343 soybean accessions acquired from the GRIN database. Traits are: seed weight (SW), Seed Length (SL), seed breadth (SB), seed height (SH), and volume.

Accession	MG	Country	SW (g/100s)	SL (mm)	SB (mm)	SH (mm)	Volume (mm^3)
PI 4243371	5	South Korea	36.62	9.35	7.03	8.56	564.93
PI 424576	5	South Korea	32.28	9.69	5.94	8.58	495.96
PI 424586	5	South Korea	31.28	9.78	5.91	8.54	496.43
PI 424593	5	South Korea	31.89	9.61	6.13	8.76	516.43
PI 458159	5	South Korea	33.31	9.58	6.51	8.30	519.51
PI 458174	5	South Korea	30.81	9.39	6.68	8.57	537.58
PI 509081	5	South Korea	30.45	9.27	6.18	8.05	462.33
PI 408254	6	South Korea	29.30	8.84	6.41	8.10	460.82
PI 424139	6	South Korea	26.76	8.73	6.16	7.59	415.64
PI 424185	6	South Korea	32.78	9.09	6.55	8.43	502.68
PI 424333	6	South Korea	21.68	8.27	4.85	7.43	297.80
PI 424375	6	South Korea	30.38	9.49	6.32	7.68	459.76
PI 424439	6	South Korea	23.82	8.82	5.35	7.66	362.04
PI 458196	6	South Korea	26.59	8.76	6.04	7.64	421.58
PI 458219	6	South Korea	33.30	9.40	6.44	8.18	495.90
PI 458245	6	South Korea	31.52	9.16	6.34	8.08	470.53
PI 458251	6	South Korea	30.36	10.41	5.81	8.08	489.45
PI 509086	6	South Korea	29.68	8.63	6.46	8.25	460.49
PI 509093	6	South Korea	33.63	9.82	6.36	8.26	516.86
PI 408051	7	South Korea	28.76	8.63	6.47	7.69	429.86
PI 458242	7	South Korea	27.21	8.57	6.31	7.62	412.65
PI 194626	00	Sweden	26.22	8.69	6.24	7.70	418.04
PI 194647	00	Sweden	19.70	8.30	5.24	7.30	316.67
PI 196486	00	Sweden	21.19	8.64	5.20	7.18	322.55
PI 518758	1	Taiwan	26.67	8.12	6.69	7.90	430.12
PI 536547B	3	Taiwan	29.03	9.30	6.11	8.27	469.10
PI 536547C	3	Taiwan	27.98	9.14	6.05	7.88	435.14
PI 561292A	3	Taiwan	24.79	8.82	6.21	7.71	422.87
PI 561288	4	Taiwan	30.30	9.05	6.19	8.13	456.34
PI 561292B	4	Taiwan	32.99	9.35	6.48	8.40	509.88
PI 561293	4	Taiwan	30.21	9.39	6.15	8.12	468.95
PI 417322	5	Unknown	35.17	11.69	5.06	9.15	540.67
FC 21340	0	Unknown	23.49	8.57	6.01	7.76	399.38
PI 567193	1	Unknown	21.75	8.53	6.40	7.89	437.17
PI 548587	3	USA	24.68	9.12	5.84	7.40	392.96
PI 548624	3	USA	28.16	8.31	6.73	8.11	453.52
PI 87165	3	USA	24.81	8.16	6.24	7.46	385.73
PI 548559	4	USA	25.94	9.33	5.69	7.73	412.99

Chapter 3

Harvest Optimization of Edamame at R6 Reproductive Stage to Achieve Maximum Pod Weight and Seed Size While Maintaining Acceptable Green Pod Color

Abstract

Edamame is a food-grade soybean (*Glycine max* (L.) Merr.) that is harvested immature between the R6 and R7 reproductive stages. To be labeled a premium product, the edamame market demands adequate pod size and green color. A staggered harvest season is critical for the commercial industry to harvest and process the crop in a timely manner. Currently, there is little information to assist in predicting the most optimum time to harvest when the pods are at their collective largest size and greenest color. The objectives of this study were to: i) estimate the optimum harvest date of three commercial edamame varieties, ii) calculate the harvest window for each variety, and iii) determine the effects of planting date and variety on pod weight and green color. Three varieties were evaluated in three staggered planting dates at two locations over two years. Pod weight and color data were analyzed to estimate the most optimum harvest date and harvest window and to determine differences between planting dates and varieties. The early maturing variety (8080) had a smaller decrease in reproductive growth period (9 days) compared to the later - maturing varieties (R08-4002 and R09-345) (21 days). The two later-maturing varieties had similar optimum days to harvest (approximately 115 days, 94 days, and 85 days for PD 1, 2, and 3, respectively), indicating consistency among varieties with similar maturity groups. Planting the three varieties over three dates allowed an extended harvest season from mid-August to early-mid October. Generally, the harvest window for each variety was very short (<5 – 7 days). Pod weight was variety specific and green pod color was planting date and variety specific. The observations of this research will help define a planting, maintenance, and harvest strategy for edamame production.

Introduction

Edamame (vegetable soybean) is a food-grade soybean (*Glycine max* (L.) Merr.), which is harvested immature between the reproductive stages of R6-R7, where the beans fill 80 – 90 percent of the pod (Konovsky et al., 1994; Shanmugasundaram and Yan, 2004). As a vegetable product, the appearance of the pod and bean must be acceptable. The main physical attributes

of edamame include large seed (>30g/100 seed at maturity) and green crescent shaped pods with two or three seed (Mentreddy et al., 2002; Shanmugasundaram and Yan, 2004).

Production of edamame in the United States is thought to have started in the 1950s including home gardens and food processers. Demand for edamame in the United States has seen a dramatic increase since the early 2000s (Mentreddy et al., 2002). Nuss (2013) reported that between 22,600 to 27,000 Megagrams of edamame per year was consumed in the U.S., estimated to be a $175 million to $200 million market.

A private company in Mulberry, Arkansas is reported to be the first major domestic producer and processer in the United States (University of Arkansas System, 2013). The United States is one of the top soybean producing countries; therefore, has the capability to produce edamame (Nuss, 2013). Producing edamame soybean is very similar to producing grain soybean (Ogles, 2016). The similarities include photoperiod sensitivity, fertilization practices, disease management, and irrigation (Ross, 2013). Machine harvest is also a similarity with commodity soybean, although a modified green bean picker is typically used for harvesting fresh edamame pods (UACES, 2013). To spread out crop risks and to even the flow of materials entering post-harvest processing, the edamame crop is typically stagger-planted through various dates and maturity-group combinations. Nolen et al. (2016) reported techniques such as these can extend the harvesting season to several months, and that a staggered harvest is critical due to the short window a variety will have acceptable seed size and color. It has been reported that the range from R5.8 – R7.0 can be 18-20 days (Purcell et al., 2014); however, Nolen et al. (2016) suggested the harvest window for an acceptable edamame product can be less than 18 days.

Soybeans will mature faster as the nights become longer (Garner and Allard, 1920). Garner and Allard (1920) added phototropism is a major factor in soybean yield. Johnson et al. (1960) indicated phototropism can affect later stages of reproductive development, not only triggering flowering. In addition, some varieties are less sensitive than others to delayed planting

and phototropism (Johnson et al., 1960); where very early varieties (maturity group (MG) 00 and 0) have been reported to not be sensitive (Polson, 1972). In addition, as the relative maturity increases, the soybean reproductive growth stages become increasingly more sensitive to long nights (Johnson et al., 1960; Major et al., 1975b).

The ability to predict the harvest time of many horticulture crops is based on accumulated thermal units throughout the crop's growing season (Baker and Reddy, 2001; Miller et al., 2001; Oliver and Annandale, 1998). Baker and Reddy (2001) calculated thermal units (Tu) as:

$$Tu = [(T_{max} + T_{min})/2] - T_b$$

where Tu is the accumulated thermal units for each day, T_{max} is the daily maximum temperature, T_{min} is the daily minimum temperature, and T_b is the base temperature below which a crop has no development. For a growing season, the daily Tu values are summed to calculate the accumulated thermal units. The base temperature in which growth and development of soybean stops is +7° Celsius (Boote et al., 1998). Previous research has suggested it is possible to use temperature in correlation with growth (Major et al., 1975b), but it has also been reported that predicting growth stages using thermal units may be no more accurate than using calendar days (Major et al., 1975a).

The first objective of this study was to evaluate the optimum harvest date among various planting dates of three different varieties and MG: 8080 (MG 3.5), R08-4002 (MG 5.8), and R09-345 (MG 6.0). The second objective was to estimate the harvest window for each variety by planting date. The third objective was to determine the effects of planting date and variety on pod weight and green color.

It was hypothesized that the optimum harvest date for each variety by planting date can be predicted by using either number of days or accumulated Tu since date of emergence (VE) and that the harvest window to maintain acceptable green color is will differ between varieties

and planting dates. It was also hypothesized the pod weight and color will differ across planting dates and varieties due to shortened vegetative growth and genetics.

Materials and Methods

Experimental Design

The experimental plan was a split-split plot design with three replications. The whole plot was planting date (PD1, PD2, and PD3), the split-plot was variety (8080, R08-4002, and R09-345), and the split-split plot was harvest date nested within planting date by variety (spaced 5 days apart from R5.8 – R7).

The experiment was conducted over two years (2014 and 2015) at two locations. The two locations were the Arkansas Agricultural Research and Extension Center (AAREC) in Fayetteville and the Vegetable Research Station in Kibler, AR. The soil was Johnsburg silt loam (fine-silty, mixed, active, mesic Aquic Fragiudults) in Fayetteville and Roxana very fine sandy loam (coarse-silty, mixed, superactive, nonacid, thermic Typic Udifluvents) in Kibler (Soil Survey Staff, 2017).

The three planting dates represented planting in middle to late May, middle to late June, and the middle of July. In both locations, the plots followed fallow ground (2014) and soybeans (2015); and the fields were cultivated to ensure the seed had adequate soil contact. The seed was planted approximately 1.9 cm deep to ensure adequate soil moisture. Each plot consisted of four, 10.7m long and 0.91m wide rows. The seeding rate was 33 seed per row meter, resulting in approximately 16 seed per row meter after germination.

Fertilizer was applied in both locations and years based on soil test from the University of Arkansas. Weeds were controlled by applying ChargerMax, Scepter 70D, and FlexStar in Fayetteville (2014 and 2015); and Dual Magnum and Pursuit in Kibler (2014 and 2015.) Stink bugs (*Nezara viridula*) were controlled in Fayetteville by applying Grizzly Z in 2014 and 2015. The irrigation schedule was based on visual observation of the soil moisture content. The plots

in Fayetteville (2014 and 2015) and Kibler (2014) were irrigated five times; however, the plots in Kibler (2015) were irrigated four times.

Approximately every five days, a total of 100 pods were randomly picked by hand within the middle two rows of each four-row plot. The harvest for each plot began approximately at the R5.8 reproductive stage and continued until the R7 reproductive stage.

Tsay and Sheu (1991) indicated that freshly harvested edamame should be cooled and then frozen to maintain the best quality; therefore, after harvest, the pods were sealed in plastic bags and placed on ice until they were stored in a refrigerator at 1.6ºC. The 100 pods were weighed (Mebrahtu et al., 1991) then immediately blanched in a 100°C. water (tap water) bath to simulate the green color of the commercially available frozen edamame product (Mozzoni et al., 2009).

Color Measurement

Color was measured with a HunterLab Color Flex (Hunter Associates Laboratory Inc., Reston, VA, U.S.A.). The instrument was calibrated with a black glass tile and a white standard tile with values of a* (-0.93) and b* (1.02). A green standard tile with values of a* (-25.30) and b* (13.71) was used to validate the calibration. The green color was interpreted as hue: calculated as (degrees(ATAN2)(a*,b*)) (Rayaprolu et al., 2015). Lawless and Heymann (1998) described hue as how close a color is to pure red, yellow, green, or blue with values of 0º, 90º, 180º, or 270º, respectively.

Thermal Unit Calculation

The daily maximum and minimum air temperature were measured by a local weather station at each location. The Tu was calculated as described by Baker and Reddy (2001) with a base temperature (+7 º C) suggested by Boote et al. (1998).

Statistical Analysis

Data were analyzed using the PROC GLIMMIX procedure of SAS 9.4 software (SAS Institute, 2014). The model information consisted of response distribution (Gaussian), link

function (identity), variance function (default), variance matrix (not blocked), estimation technique (restricted maximum likelihood), degrees of freedom method (Kenward-Roger), and fixed effects SE adjustment (Kenward-Roger). The fixed effects in the PROC GLIMMIX model were: planting date (whole plot), variety (split-plot), harvest date nested within planting date and variety (split-split plot) and the interactions of planting date by variety. The random effects were year, location, and the interactions of year by location, block nested within year by location, and planting date by block nested within year by location. Year and location did not initially have a significant effect on pod weight or hue (alpha = 0.05); therefore, were fitted as random effects.

Results

Planting Date and Growth Stages

The planting date code (PD code), planting date (PD), date of emergence (VE), and when the plots began the first reproductive stage (R1) (referring to onset of flowering) are presented in (Table 1). The planting date code (PD code) will represent the planting date in subsequent tables.

The Optimum Harvest Date

Variety 8080 (MG 3.5)

The range of number of days between VE and R1 for all three planting dates was from 23 to 32 days for both 2014 and 2015, even though the date of emergence ranged from 30 May to 22 July. The most number of days (32) between VE and R1 was in the second planting date of Fayetteville with an emergence date of 16 June 2014. The emergence date of 12 July 2014 resulted in 29 days between VE and R1 at Fayetteville and Kibler, where the earliest emergence date (30 May 2014) resulted in 27 days between VE and R1. The third planting date in Fayetteville was the exception, where an emergence date of 22 July 2015 resulted in 23 days between VE and R1 (Table 1).

The average date of VE, across all locations and years, for planting dates one, two, and three were 2 June, 24 June, and 15 July, respectively, and the number of days between VE and

harvest date were 83, 82, and 71 for PD 1, PD 2, and PD 3, respectively (Table 2). The average accumulated Tu between VE and HAR was 1546, 1515, and 1268 for PD 1, PD 2, and PD 3, respectively. The average daily Tu for each planting date were: 18.63 (PD 1), 18.48 (PD 2), and 17.86 (PD 3). The average pod weight for the three planting dates were: 293.43 g/100 pods for PD 1, 258.19 g/100 pods for PD 2, and 305.51 g/100 pods for PD 3. The average hue for the three planting dates were: 111.13 º for PD 1, 110.64 º for PD 2, and 110.70 º for PD 3 (Table 2).

Variety R09-345 (MG 6.0)

The range of number of days between VE and R1 was from 33 (emergence date of 22 July 2015) to 54 (emergence date of 26 May 2015). The average number of days between VE and R1 for PD 1, PD 2, and PD 3 were 51, 43, and 38.7, respectively, with an average planting date of 28 May, 24 June, and 15 July.

The average date of VE, across all locations and years, for PD 1, PD 2, and PD 3 were 28 May, 24 June, and 15 July, respectively; and the number of days between VE and HAR were 112, 92, and 85 for PD 1, PD 2, and PD3, respectively (Table 2). The average accumulated Tu between VE and HAR was 2037, 1666, and 1446 for PD 1, PD, 2, and PD 3, respectively. The average daily Tu for each planting date were: 18.19 (PD 1), 18.11 (PD 2), and 17.01 (PD 3). The average pod weight for the three planting dates were 149.60 g/100 pods, 161.23 g/100 pods, and 190.00 g/100pods for PD 1, PD 2, and PD 3, respectively. The average hue for the three planting dates were: 110.17º, 110.03º, and 110.04º for PD 1, PD 2, and PD 3, respectively (Table 2).

Variety R08-4002 (MG 5.8)

The range of number of days between VE and R1 ranged from 31 (emergence date of 22 July 2015) to 52 (emergence date of 26 May 2014). The average number of days between VE and R1 for PD 1, PD 2, and PD 3 were 49.5, 39.3, and 35.7 days, respectively; with an average planting date of 28 May (PD 1), 24 June (PD 2), and 15 July (PD 3). In all three

planting dates, the number of days between VE and R1 was lower for every plot with a later emergence date (Table 1).

The average date of VE, across all locations and years, for PD 1, PD 2, and PD 3 were 28 May, 24 June, and 15 July, respectively; and the number of days between VE and HAR were 118, 96, and 85 for PD 1, PD 2, and PD 3, respectively (Table 2). The average accumulated Tu between VE and harvest were 2126, 1723, and 1446 for PD 1, PD 2, and PD 3, respectively. The average daily Tu for each planting date were 18.02, 17.95, and 17.01 for PD 1, PD 2, and PD 3, respectively. The average pod weight for the three planting dates were 152.53 g/100 pods, 143.91 g/100 pods, and 150.93 g/100 pods, for PD 1, PD 2, and PD 3, respectively. The average hue for the three planting dates were 110.21º, 111.65º, and 113.43º for PD 1, PD 2, and PD 3, respectively (Table 2).

Harvest Window

In addition to the optimum harvest date, a range of dates of acceptable harvest was determined using the harvest dates (split-split plot) for each planting date by variety (whole plot by split-plot) with location and year as random variables. Within the harvest window, the pod weight and hue did not differ ($p > 0.05$) than that for the optimum harvest (largest pod weight and highest hue combination) (Table 3).

The yearly harvest windows, across all three planting dates, were 17 Aug. – 24 Sept. for variety 8080 and 17 Sept. – 08 Oct. for varieties R09-345 and R08-4002. The confirmed harvest window for 8080 in PD 1 was 7 d and was 5 d for PD 2 and PD 3. There was no confirmed multiple day harvest window in all three planting dates for R09-345 as there were no multiple harvest dates with a pod weight or green color similar to the most optimum harvest date. The variety R08-4002 had a confirmed harvest window of 4 d in PD 2; however, there was not a confirmed harvest window in PD 1 and PD 3 (Table 3).

Differences Among Planting Dates and Varieties

Pod Weight

Pod weight differed among all three varieties ($p<0.0001$) (Table 4). The varieties with the largest to smallest pod weight were: 8080 (259.52 g/100 pods), R09-345 (165.53 g/100 pods), and R08-4002 (134.05 g/100 pods) (Table 5).

Hue

Hue differed among the factors planting date and variety ($p<0.05$) (Table 6). Planting date 1 (110.4º) and PD 2 (109.34 º) had a hue significantly closer to pure green ($p<0.05$) than PD 2 (106.64 º) across all years, locations, and varieties. The varieties R08-4002 (109.52 º) and 8080 (109.27 º) had a hue closer to pure green than R09-345 (107.6 º) across all years, locations, and planting dates (Table 7).

Discussion

Proper irrigation practices and pesticide control are essential at the beginning of the reproductive stages (R1). Adequate soil moisture prior to full bloom is essential to reach full yield and seed-size potential. In addition, scouting for and controlling pest becomes increasingly more critical beginning at the R1 reproductive growth stage (Ashley and Ethridge, 1973; Faske et al., 2014; Lorenz et al., 2000; Tacker and Vories, 2000). The date each plot reached the R1 reproductive growth stage was recorded which can be used to schedule crop maintenance.

Optimum Harvest Date

When planting is staggered from the middle of May to the middle of July, the harvest season for 8080 (MG 3.5) can be extended from the middle of August until late September. With the same staggered planting dates, the harvest season for R08-4002 (MG 5.8) and R09-345 (MG 6.0) can be extended from the middle of September until early-mid October. The results of this research suggest a harvest season spanning from mid-August to early-mid October is possible using varieties with different MG and staggered planting dates (MG 3.5, 5.8, and 6.0) (PD mid-May, mid-June, and mid-July). Nolen et al. (2016) reported an edamame harvest

season can be from early July until late October using four production methods: high tunnel, field/plastic, field/no plastic, and conventional. The harvest window of this experiment only differed in the early July harvest reported by Nolen et al. (2016). Further research is needed to determine when edamame varieties of different MG would be ready to harvest when planting in late April or early May; however, caution should be taken to avoid an early or late frost (Daly et al., 2010).

It has been reported that soybean varieties with earlier MG tend to be less sensitive to later planting dates, indicating, between earlier and later planting dates, the number of days between VE and R1 and/or days between VE and HAR at the R6 reproductive stage will have a smaller decrease (Johnson et al., 1960; Criswell et al., 1972; Polson, 1972; Major et al., 1975b; Salmeron et al., 2014). The differential timing between VE and HAR between 8080 (MG 3.5) and R08-4002 (MG 5.8) / R09-345 (MG 6.0) supports the previous findings. Although all three varieties had a decrease in the number of days of vegetative and reproductive growth, the variety 8080 (MG 3.5) had a smaller decrease than the varieties R08-4002 (MG 5.8), and R09-345 (MG 6.0). The average decrease in the number of days between VE to R1 from PD 1 to PD 3 was 0.5, 13.8, and 12.3 for 8080, R08-4002, and R09-345, respectively (Table 1) and the average decrease in number of days between R1 to harvest was 12.5, 19.8, and 25.7 for 8080, R08-4002, and R09-345, respectively (Table 1 and Supplemental Table 1). The two later-maturing varieties had a decrease in the number of days between R1 and harvest for each subsequent planting date, where 8080 only decreased in the number of days from planting dates two to three. All three varieties were more sensitive to later planting dates between the stages of R1 to harvest than between VE-R1. The results from Setiyono et al. (2007) supported the observations of this research, where the number of days between VE to R1 and R1 to HAR may not have similar correlations to either the number of days or accumulated thermal units. There was no evidence of a specific accumulated thermal unit to predict the day of flowering or

day of harvest, as has been reported for other crops (Baker et al., 2001; Miller et al., 2001; Oliver and Annandale, 1998).

Harvest Window

The variety 8080 had a harvest window of seven days for planting date one and five days for planting dates two and three, where R08-4002 had a harvest window of four days in planting date two. The varieties R09-345 and R08-4002 did not have a confirmed extended harvest window in PD 1-3, and PD 1 and 3, respectively. These results signify there were no two harvest within a planting date that were similar to the most optimum harvest. Since the average harvest was spaced five days apart, the optimum harvest can be interpreted as less than 10 days (<5 days before and < 5 days after).

Differences Among Planting Dates and Varieties

Pod Weight

Panthee et al. (2004) suggested seed size has high heritability, indicating the pod weight should be controlled more by genetic than environmental variances (Panthee et al., 2004). Beatty et al. (1982) noted that seed weight did not differ from a 15 April to 15 May planting date, but decreased significantly each month from a 15 May to 15 July planting date. In this study, there were no differences in pod weight across planting dates; however, in contrast to Beatty et al. (1982), the third planting date had a larger pod weight (191.49 g/100 pods) than the second (184.15 g/100 pods) or first (183.46 g/100 pods) planting dates.

Similar to SW at the R8 growth stage (mature and dry), there were significant differences in pod weight ($p<0.0001$) across varieties at the R6 reproductive growth stage. From largest to smallest, the pod weights were: 259.52 g/100 seed (8080), 165.53 g/100 seed (R09-345), and 134.05 g/100 seed (R08-4002). These values correlate with the historical R8 (mature) seed size (g/100 seed) measured by the University of Arkansas Soybean Breeding Program. The values of the mature seed are: 33 g/100 seed (8080), 26.5 g/100 seed (R09-345), and 23.5 g/100 seed (R08-4002) (P. Chen, personal communication, July 2017).

Hue

The hue value of the first and third planting dates were closer to pure green than the second planting date (Table 7). All three varieties were exposed to the lowest average air temperatures in the third planting date (Table 2 and Figure 1). The lower average temperatures in the third planting date may have resulted in a lower respiration rate resulting in an increase in photosynthate material (Atkin and Tjoelker, 2003). This may help explain why the pod weights were slightly greater in the third planting date than the first and second.

Although R09-345 (black seed color at R8 reproductive growth stage) is green at R6, the pods have a lower hue than the other two varieties (Table 7). This result suggest varieties that turn black or brown will not have the same hue at R6 compared to varieties that either stay green or turn yellow at the R8 reproductive growth stage.

Conclusions

Due to sensitivity of the length of night (dark) hours, varieties of similar MG should flower and be ready for R6 harvest at approximately the same time each year; therefore, results of this research should help predict growth stages and time of harvest. An edamame harvest season spanning the months of mid-August to early-mid October is possible due to multiple MG and staggered planting dates. The results of this research suggest varieties with early MG may be less sensitive to later planting dates. The lower sensitivity to late planting dates may enable varieties with earlier MG to lose less potential yield then later-maturing varieties. This information is helpful for edamame farmers wanting to grow two crops per year. Pod weight and green color are important characteristics for edamame. The results of this study indicate the choice of variety is important for both variables.

References

Ashley, D.A., and W.J. Ethridge. 1973. Irrigation effects on vegetative and reproductive development of three soybean varieties. Agron. J. 70:467-471.

Atkin, O.K., and M.G. Tjoelker. 2003. Thermal acclimation and the dynamic response of plant respiration to temperature. Trends Plant Sci. 8:343-351.

Baker, J.T., D.I. Leskovar, V.R. Reddy, and F.J. Dainello. 2001. A Simple phenological model of muskmelon development. Ann. Bot. 87:615-621.

Baker, J.T., and V.R. Reddy. 2001. Temperature effects on phenological development and yield of muskmelon. Ann. Bot. 87:605-613.

Beatty, K.D., I.L. Eldridge, and A.M. Simpson, Jr. 1982. Soybean response to different planting patterns and dates. Agron. J. 74:859-862.

Boote, K.J., J.W. Jones, and G. Hoogenboom. 1998. Simulation of crop growth: CROPGRO model. p. 651-692. In R.M. Peart and R.B. Curry (ed.) Agricultural systems modeling and simulation. Marcel Dekker, New York.

Criswell, J.G., and D.J. Hume. 1972. Variation in sensitivity to photoperiod among early maturing soybean strains. Crop Sci. 12:657-660.

Daly, C., M.D. Halbleib, J.I. Smith, and W.P. Gibson. 2010. Development of a new USDA plant hardiness zone map for the United States. J. Appl. Meteorol. Climatology. 51:242-264.

Faske, T., T. Kirkpatrick, J. Zhou, and I. Tzanetakis. 2014. Soybean diseases, pp. 1-18. *In* Arkansas Soybean Handbook, MP197. Ark. Coop. Ext. Ser., Little Rock, AR.

Garner, W.W., and H.A. Allard. 1920. Effect of the relative length of day and night and other factors of the environment on growth and reproduction in plants. J. Agr. Res. 18:553-606.

Johnson, H.W., H.A. Borthwick, and R.C. Leffel. 1960. Effects of photoperiod and time of planting on rates of development of the soybean in various stages of the life cycle. Bot. Gaz. 122:77-95.

Konovsky, J., T.A. Lumpkin, and D. McClary. 1994. Edamame: the vegetable soybean. Pages 173-81 in A.D. O'Rourke (ed), Understanding the Japanese food and agrimarket: a multifaceted opportunity. Haworth Press, Binghamton.

Lawless, H.T., and H. Heymann. 1998. Sensory evaluation of food. Chapman and Hall, New York. p. 819.

Lorenz, G., D. Johnson, G. Studebaker, C. Allen, and S. Young, III. 2000. Insect pest management in soybeans, pp. 1-11. *In* Arkansas Soybean Handbook, MP 197. Ark. Coop. Ext. Ser., Little Rock, AR.

Major, D.J., D.R. Johnson, and V.D. Luedders. 1975a. Evaluation of eleven thermal unit methods for predicting soybean development. Crop Sci. 15:172-174.

Major, D.J., D.R. Johnson, J.W. Tanner, and I.C. Anderson. 1975b. Effects of day length and temperature on soybean development. Crop Sci. 15:174-179.

Mebrahtu, T., A. Mohamed, and W. Mersie. 1991. Green pod yield and architectural traits of selected vegetable soybean genotypes. J. Prod. Agric. 4:395-399.

Mentreddy, S.R., A.I. Mohamed, N. Joshee, A.K. Yadav. 2002. Edamame: A nutritious vegetable crop. P. 432-438. In: J. Janick and A. Whipkey (eds.), Trends in new crops and new uses. ASHS Press, Alexandria, VA.

Miller, P., W. Lanier, and S. Brandt. 2001. Using growing degree days to predict plant stages. Montana State University Extension Service.

Mozzoni, L.A., R.O. Morawicki, and P. Chen. 2009. Canning of vegetable soybean: procedures and quality evaluations. Int. J. Food Sci.Technol. 44:1125-1130.

Nolen, S., B. Zhang, and M.K. Kering. 2016. Increasing fresh edamame bean supply through season extension techniques. J. Hortic. 3:170. doi:10.4172/2376-0354.1000170.

Nuss, J. 2013. US soybean farmers see growth potential in edamame. CNSnews.com. http://www.cnsnews.com/news/article/us-soybean-farmers-see-growth-potential-edamame (accessed 11 Jan. 2017).

Ogles, C.Z., E.A. Guertral, and D.B. Weaver. 2016. Edamame cultivar evaluation in central Alabama. Agron. J. 108:2371-2378.

Oliver, F.C., and J.G. Annandale. 1998. Thermal time requirements for the development of green pea (*Pisum sativum* L.). Field Crops Res. 56:301-307.

Panthee, D.R., V.R. Pantalone, D.R. West, A.M. Saxton, and C.E. Sams. 2004. Quantitative trait loci for seed protein and oil concentration, and seed size in soybean. Crop Sci. 45:2015-2022.

Polson, D.E. 1972. Day-neutrality in soybeans. Crop Sci. 12:773-776.

Purcell, L.C., M. Salmeron, and L. Ashlock. 2014. Soybean growth and development, pp. 1-8. *In* Arkansas Soybean Handbook, MP 197. Ark. Coop. Ext. Ser., Little Rock, AR.

Rayaprolu, S., N. Hettiarachchy, M. Aldoury, S. Cho, D. Moseley, and P. Chen. 2015. Physical and textural attributes of freeze-dried genetically modified and non-genetically modified soy beans. J. Food Nutr. Sci. 3:119-125.

Ross, J. 2013. Edamame production practices: What we know now. Univ. of Arkansas. https://www.uaex.edu/farm-ranch/crops-commercial-horticulture/edamame/docs/Edamame%20-%20What%20We%20Know%20Now.pdf.

Salmerón, M., E.E. Gbur, F.M. Bourland, N.W. Buehring, L. Earnest, F.B. Fritschi, et al. 2014. Soybean maturity group choices for early and late plantings in the midsouth. Agron J. 106:1893-1901.

SAS Institute. 2014. The SAS System for Windows. Version 9.4. SAS Inst. Inc., Cary, NC.

Setiyono, T.D., A. Weiss, J. Specht, A.M. Bastidas, K.G. Cassman, and A. Dobermann. 2007. Understanding and modeling the effect of temperature and day length on soybean phenology under high-yield conditions. Field Crops Res. 100:257-271.

Shanmugasundaram, S., and M.R. Yan. 2004. Global expansion of high value vegetable soybean. VII world soybean conference/IV international soybean processing and utilization conference/III Brazilian Soybean Congress, 29 Feb–5 Mar 2004. Foz do Iguassu, Brazil 915–920.

Soil Survey Staff, Natural Resources Conservation Service, United States Department of Agriculture. Web Soil Survey. Available online at the following link: https://websoilsurvey.sc.egov.usda.gov/. Accessed [12/11/2017].

Tacker, P., and E. Vories. 2000. Irrigation, pp. 1-8. *In* Arkansas Soybean Handbook, MP197. Ark. Coop. Ext. Ser., Little Rock, AR.

Tsay L., and S. Sheu. 1991. Studies on the effects of cold storage and precooling on the quality of vegetable soybean. *In*: Shanmungasundaram, S. (ed.) Vegetable soybean: research needs for production and quality improvement. Proceeding of workshop, Kenting, Taiwan, Asian Vegetable Research and Development Center.

University of Arkansas Cooperative Extension Service. 2013. Edamame production facts. University of Arkansas Division of Agriculture. https://www.uaex.edu/farm-ranch/crops-commercial-horticulture/edamame/docs/EdamameFacts1.pdf.

Table 1. The emergence stage (VE) and beginning bloom (R1) date.

A. 8080

Year	Location	PD Code[d]	PD[c]	VE[b]	R1[a]
2014	Fayetteville	1	05/24/14	06/04/14	07/02/14
2014	Fayetteville	2	06/11/14	06/16/14	07/18/14
2014	Fayetteville	3	07/07/14	07/12/14	08/10/14
2014	Kibler	1	05/22/14	05/30/14	06/26/14
2014	Kibler	2	06/13/14	06/18/14	07/16/14
2014	Kibler	3	07/07/14	07/12/14	08/10/14
2015	Fayetteville	2	06/26/15	07/01/15	07/27/15
2015	Fayetteville	3	07/17/15	07/22/15	08/14/15
2015	Kibler	2	06/25/15	06/30/15	07/27/15

B. R08-4002

Year	Location	PD Code[d]	PD[c]	VE[b]	R1[a]
2014	Fayetteville	1	05/16/14	05/26/14	07/17/14
2014	Fayetteville	2	06/11/14	06/16/14	08/04/14
2014	Fayetteville	3	07/07/14	07/12/14	08/21/14
2014	Kibler	1	05/22/14	05/29/14	07/15/14
2014	Kibler	2	06/13/14	06/18/14	07/31/14
2014	Kibler	3	07/07/14	07/12/14	08/17/14
2015	Fayetteville	2	06/26/15	07/01/15	08/02/15
2015	Fayetteville	3	07/17/15	07/22/15	08/22/15
2015	Kibler	2	06/25/15	06/30/15	08/02/15

C. R09-345

Year	Location	PD Code[d]	PD[c]	VE[b]	R1[a]
2014	Fayetteville	1	05/16/14	05/26/14	07/19/14
2014	Fayetteville	2	06/11/14	06/16/14	08/07/14
2014	Fayetteville	3	07/07/14	07/12/14	08/22/14
2014	Kibler	1	05/22/14	05/29/14	07/16/14
2014	Kibler	2	06/13/14	06/18/14	08/03/14
2014	Kibler	3	07/07/14	07/12/14	08/23/14
2015	Fayetteville	2	06/26/15	07/01/15	08/06/15
2015	Fayetteville	3	07/17/15	07/22/15	08/24/15
2015	Kibler	2	06/25/15	06/30/15	08/07/15

[a] The onset of flowering
[b] The date of emergence
[c] The planting date for each plot
[d] The planting date code for each plot

Table 2. Emergence, harvest, pod weight, and color data for each planting date.

A. 8080

PD[z]	VE[y]	RVE[x]	HAR[w]	Avg. HD[v]	Range HD[u]	Avg. Days[t]	Avg. Tu[s]	Rang Tu[r]	Pod Weight[q]	Hue[p]
									g/100 pods	degrees
1	6/02	5/30-6/04	5	8/24	8/20-8/27	83	1546	1422-1704	293.43ab	111.13abc
2	6/24	6/16-7/01	4	9/14	9/01-9/27	82	1515	1412-1641	258.19cd	110.64bc
3	7/15	7/12-7/22	4	9/24	9/21-9/27	71	1268	1182-1368	305.51a	110.7bc

B. R09-345

PD[z]	VE[y]	RVE[x]	HAR[w]	Avg. HD[v]	Range HD[u]	Avg. Days[t]	Avg. Tu[s]	Rang Tu[r]	Pod Weight[q]	Hue[p]
									g/100 pods	degrees
1	5/28	5/26-5/29	3	9/17	9/14-9/19	112	2037	1890-2235	149.60efgh	110.17bc
2	6/24	6/16-7/01	3	9/24	9/21-9/26	92	1666	1548-1805	161.23cdef	110.03c
3	7/15	7/12-7/22	3	10/08	10/06-10/10	85	1446	1358-1556	190.00ab	110.04c

C. R08-4002

PD[z]	VE[y]	RVE[x]	HAR[w]	Avg. HD[v]	Range HD[u]	Avg. Days[t]	Avg. Tu[s]	Rang Tu[r]	Pod Weight[q]	Hue[p]
									g/100 pods	degrees
1	5/28	5/26-5/29	4	9/23	9/19-9/26	118	2126	1972-2330	152.53defgh	110.21bc
2	6/24	6/16-7/01	4	9/28	9/26-9/30	96	1723	1602-1869	143.91fgh	111.65abc
3	7/15	7/12-7/22	3	10/08	10/6-10/10	85	1446	1358-1556	150.93efgh	113.43abc

Means following the same letter are not significantly different ($p>0.05$) (for pod weight: means are separated within 8080 and means are separated among R09-345 and R08-4002; for hue: means are separated among all three varieties.)

[z] The planting date code (whole plot factor)

[y] The average date the plants emerged within each planting date (both locations and years included)

[x] The range of dates of emergence (both locations and years included)

[w] The harvest code (split-split plot factor)

[v] The average harvest date within each planting date (both locations and years included)

[u] The range of harvest dates (both locations and years included)

[t] The average number of days from VE to harvest (both locations and years included)

[s] The average number of thermal heat units from VE to harvest (both locations and years included)

[r] The range of thermal heat units from VE to harvest (both locations and years included)

[q] The average pod weight for each planting date (both locations and years included)

[p] The average hue for each planting date (both locations and years included) (higher hue value – more green color)

Table 3 Harvest Window for each planting date.

A. 8080

PD Code[u]	HAR. Code[t]	Avg. Har. Date[s]	Har. Window[r]	Pod Weight[q]	Hue[p]
			days	g/100 pods	degrees
1	3	8/17		275.68b	111.90abc
1	4	8/21	7	279.12b	112.01abc
1	5*	8/24		293.43ab	111.13abc
2	3	9/9	5	245.74d	112.11abc
2	4*	9/14		258.19cd	110.64bc
3	3	9/19	5	284.08ab	114.70abc
3	4*	9/24		305.51a	110.70bc

B. R09-345

PD Code[u]	HAR. Code[t]	Avg. Har. Date[s]	Har. Window[r]	Pod Weight[q]	Hue[p]
			days	g/100 pods	degrees
1	3*	9/17	0	149.60efgh	110.17bc
2	3*	9/24	0	161.23cdef	110.03c
3	3*	10/08	0	190.00ab	110.04c

C. R08-4002

PD Code[u]	HAR. Code[t]	Avg. Har. Date[s]	Har. Window[r]	Pod Weight[q]	Hue[p]
			days	g/100 pods	degrees
1	4*	9/23	0	152.53defgh	110.21bc
2	3	9/24	4	137.09gh	113.55abc
2	4*	9/28		143.91fgh	111.65abc
3	3*	10/08	0	150.93efgh	113.43abc

Means with the same following letter are not significantly different (p>0.05)
(for pod weight: means are separated within 8080 and means are separated among R09-345 and R08-4002; for hue: means are separated among all three varieties.)
* The optimum harvest date
[u] The planting date code (whole plot factor)
[t] The harvest code (split-split plot factor)
[s] The average harvest date within each planting date (both locations and years included)
[r] The confirmed number of days the variety in a given planting date can be harvested with optimum quality
[q] The average pod weight for each planting date (both locations and years included)
[p] The average hue for each planting date (both locations and years included) (higher hue value = more green color)

Table 4 Type III test of fixed effects for pod weight.

A. Fixed effects

Plot	DF	EDF	F Value	Pr > F
Planting Date	2	23	2.01	0.1574
Variety	2	355	902.06	<0.0001

Table 5. Separation of pod weight estimate for variety.

Variety	Estimate	SE	DF	t Value	Pr > |t|
8080	259.52a*	5.8296	1.20	44.52	0.0069
R09-345	165.53b	5.8885	1.25	28.11	0.0104
R08-4002	134.05c	5.8885	1.25	22.77	0.0135

*Means with the same following letter are not significantly different (p>0.05)

Table 6. Type III test of fixed effects for hue.

A. Fixed effects

Plot	DF	EDF	F Value	Pr > F
Planting Date	2	36.24	8.3	0.0011
Variety	2	345.4	5.71	0.0036

Table 7. Separation of hue estimate for planting date and variety.

A. Planting date (PD)

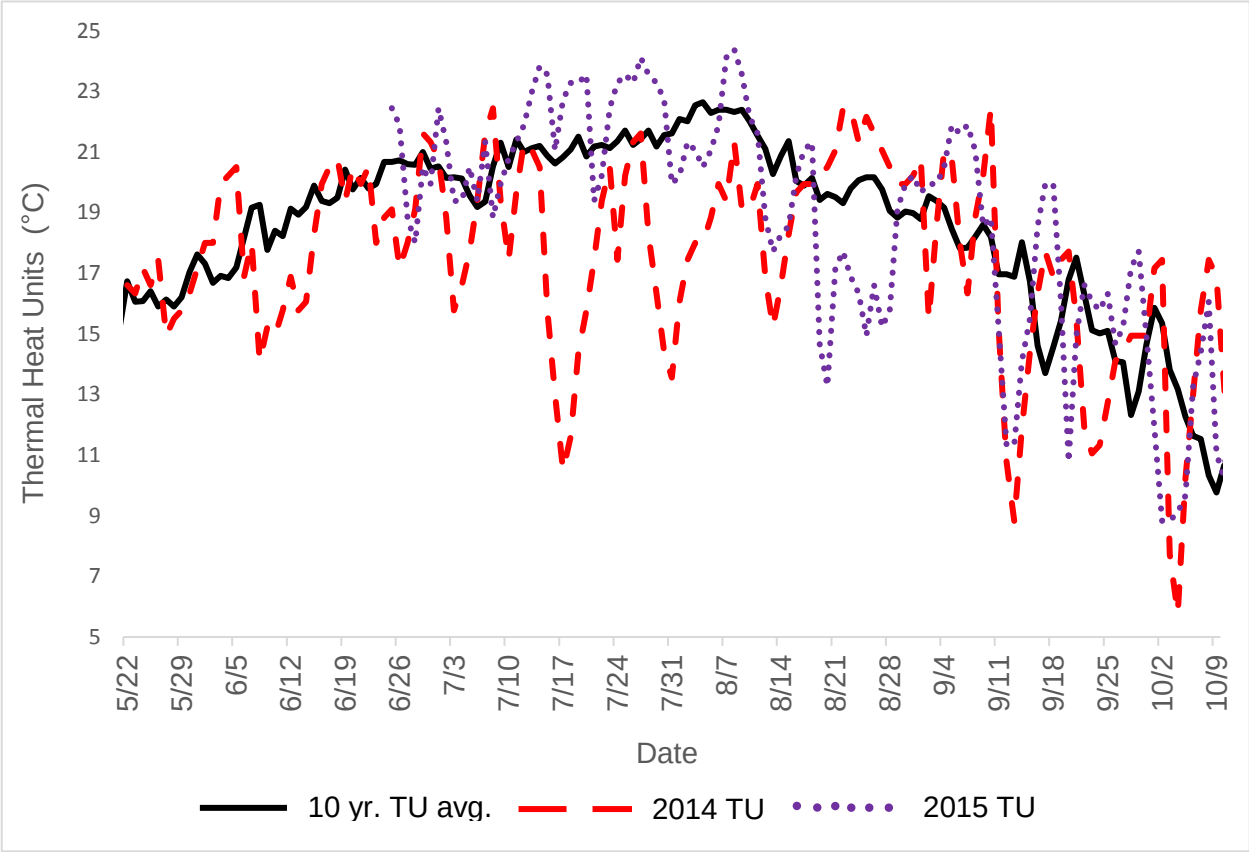
PD code	Estimate	SE	DF	t Value	Pr > |t|
1	110.40a*	1.0659	1.939	103.57	0.0001
3	109.34a	0.9726	1.564	112.42	0.0005
2	106.64b	0.9082	1.258	117.42	0.0017

B. Variety

Variety	Estimate	SE	DF	t Value	Pr > |t|
R08-4002	109.52a*	0.92	1.32	118.87	0.0013
8080	109.27a	0.90	1.186	121.9	0.0022
R09-345	107.6b	0.92	1.324	116.67	0.0013

*Means with the same following letter are not significantly different ($p>0.05$)

Figure 1. Thermal heat units for 10-year (2005 – 2015) average vs. 2014 and 2015 (Kibler).



Supplemental Table 1. Harvest date for each variety by planting date.

Year	Location	PD[b]	Variety	Har. Code[a]	Harvest Date
2014	FAY	1	8080	1	08/12/14
2014	FAY	1	8080	2	08/17/14
2014	FAY	1	8080	3	08/20/14
2014	FAY	1	8080	4	08/24/14
2014	FAY	1	8080	5	08/27/14
2014	FAY	1	8080	6	08/30/14
2014	FAY	2	8080	1	09/09/14
2014	FAY	2	8080	2	09/15/14
2014	FAY	2	8080	3	09/19/19
2014	FAY	2	8080	4	09/27/14
2014	FAY	2	8080	5	09/30/14
2014	FAY	3	8080	1	09/09/14
2014	FAY	3	8080	2	09/15/14
2014	FAY	3	8080	3	09/19/14
2014	FAY	3	8080	4	09/27/14
2014	FAY	3	8080	5	09/30/14
2014	KIB	1	8080	1	08/07/14
2014	KIB	1	8080	2	08/11/14
2014	KIB	1	8080	3	08/14/14
2014	KIB	1	8080	4	08/17/14
2014	KIB	1	8080	5	08/20/14
2014	KIB	1	8080	6	08/24/14
2014	KIB	2	8080	1	08/24/14
2014	KIB	2	8080	2	08/27/14
2014	KIB	2	8080	3	08/30/14
2014	KIB	2	8080	4	09/01/14
2014	KIB	2	8080	5	09/05/14
2014	KIB	2	8080	6	09/14/14
2014	KIB	3	8080	1	09/09/14
2014	KIB	3	8080	2	09/14/14
2014	KIB	3	8080	3	09/19/14
2014	KIB	3	8080	4	09/24/14
2014	KIB	3	8080	5	09/30/14
2014	KIB	3	8080	6	10/06/14
2014	FAY	1	R08-4002	1	09/09/14
2014	FAY	1	R08-4002	2	09/15/14
2014	FAY	1	R08-4002	3	09/19/14
2014	FAY	1	R08-4002	4	09/26/14
2014	FAY	1	R08-4002	5	09/30/14
2014	FAY	2	R08-4002	1	09/15/14
2014	FAY	2	R08-4002	2	09/19/14
2014	FAY	2	R08-4002	3	09/26/14
2014	FAY	2	R08-4002	4	09/30/14
2014	FAY	3	R08-4002	1	09/27/14
2014	FAY	3	R08-4002	2	09/30/14
2014	FAY	3	R08-4002	3	10/07/14
2014	FAY	3	R08-4002	4	10/13/14

Supplemental table 1 cont. Harvest date for each variety by planting date

2014	KIB	1	R08-4002	1	09/05/14
2014	KIB	1	R08-4002	2	09/09/14
2014	KIB	1	R08-4002	3	09/14/14
2014	KIB	1	R08-4002	4	09/19/14
2014	KIB	1	R08-4002	5	09/24/14
2014	KIB	1	R08-4002	6	09/30/14
2014	KIB	2	R08-4002	1	09/14/14
2014	KIB	2	R08-4002	2	09/19/14
2014	KIB	2	R08-4002	3	09/24/14
2014	KIB	2	R08-4002	4	09/30/14
2014	KIB	2	R08-4002	5	10/06/14
2014	KIB	3	R08-4002	1	09/24/14
2014	KIB	3	R08-4002	2	09/30/14
2014	KIB	3	R08-4002	3	10/06/14
2014	KIB	3	R08-4002	4	10/15/14
2014	FAY	1	R09-345	1	09/09/14
2014	FAY	1	R09-345	2	09/15/14
2014	FAY	1	R09-345	3	09/19/14
2014	FAY	1	R09-345	4	09/26/14
2014	FAY	1	R09-345	5	09/30/14
2014	FAY	2	R09-345	1	09/15/14
2014	FAY	2	R09-345	2	09/19/14
2014	FAY	2	R09-345	3	09/26/14
2014	FAY	2	R09-345	4	09/30/14
2014	FAY	3	R09-345	1	09/27/14
2014	FAY	3	R09-345	2	09/30/14
2014	FAY	3	R09-345	3	10/07/14
2014	FAY	3	R09-345	4	10/13/14
2014	KIB	1	R09-345	1	09/05/14
2014	KIB	1	R09-345	2	09/09/14
2014	KIB	1	R09-345	3	09/14/14
2014	KIB	1	R09-345	4	09/19/14
2014	KIB	1	R09-345	5	09/24/14
2014	KIB	1	R09-345	6	09/30/14
2014	KIB	2	R09-345	1	09/14/14
2014	KIB	2	R09-345	2	09/19/14
2014	KIB	2	R09-345	3	09/24/14
2014	KIB	2	R09-345	4	09/30/14
2014	KIB	2	R09-345	5	10/06/14
2014	KIB	3	R09-345	1	09/24/14
2014	KIB	3	R09-345	2	09/30/14
2014	KIB	3	R09-345	3	10/06/14
2014	KIB	3	R09-345	4	10/15/14
2015	FAY	2	8080	1	09/03/15
2015	FAY	2	8080	2	09/07/15
2015	FAY	2	8080	3	09/10/15
2015	FAY	2	8080	4	09/15/15

Supplemental table 1 cont. Harvest date for each variety by planting date

2015	FAY	2	8080	5	09/18/15
2015	FAY	2	8080	6	09/21/15
2015	FAY	2	8080	7	09/26/15
2015	FAY	3	8080	1	09/10/15
2015	FAY	3	8080	2	09/15/15
2015	FAY	3	8080	3	09/18/15
2015	FAY	3	8080	4	09/21/15
2015	FAY	3	8080	5	09/26/15
2015	FAY	3	8080	6	09/29/15
2015	FAY	3	8080	7	10/03/15
2015	KIB	2	8080	1	09/02/15
2015	KIB	2	8080	2	09/06/15
2015	KIB	2	8080	3	09/09/15
2015	KIB	2	8080	4	09/13/15
2015	KIB	2	8080	5	09/18/15
2015	KIB	2	8080	6	09/24/15
2015	KIB	2	8080	7	09/27/15
2015	FAY	2	R08-4002	1	09/15/15
2015	FAY	2	R08-4002	2	09/18/15
2015	FAY	2	R08-4002	3	09/21/15
2015	FAY	2	R08-4002	4	09/26/15
2015	FAY	2	R08-4002	5	09/29/15
2015	FAY	2	R08-4002	6	10/03/15
2015	FAY	2	R08-4002	7	10/10/15
2015	FAY	3	R08-4002	1	09/26/15
2015	FAY	3	R08-4002	2	10/03/15
2015	FAY	3	R08-4002	3	10/10/15
2015	FAY	3	R08-4002	4	10/17/15
2015	KIB	2	R08-4002	1	09/13/15
2015	KIB	2	R08-4002	2	09/18/15
2015	KIB	2	R08-4002	3	09/24/15
2015	KIB	2	R08-4002	4	09/27/15
2015	KIB	2	R08-4002	5	09/30/15
2015	KIB	2	R08-4002	6	10/02/15
2015	KIB	2	R08-4002	7	10/04/15
2015	KIB	2	R08-4002	8	10/11/15
2015	FAY	2	R09-345	1	09/15/15
2015	FAY	2	R09-345	2	09/18/15
2015	FAY	2	R09-345	3	09/21/15
2015	FAY	2	R09-345	4	09/26/15
2015	FAY	2	R09-345	5	09/29/15
2015	FAY	2	R09-345	6	10/03/15
2015	FAY	2	R09-345	7	10/10/15
2015	FAY	3	R09-345	1	09/26/15
2015	FAY	3	R09-345	2	10/03/15
2015	FAY	3	R09-345	3	10/10/15
2015	FAY	3	R09-345	4	10/17/15

Supplemental table 1 cont. Harvest date for each variety by planting date

2015	KIB	2	R09-345	1	09/13/15
2015	KIB	2	R09-345	2	09/18/15
2015	KIB	2	R09-345	3	09/24/15
2015	KIB	2	R09-345	4	09/27/15
2015	KIB	2	R09-345	5	09/30/15
2015	KIB	2	R09-345	6	10/02/15
2015	KIB	2	R09-345	7	10/04/15
2015	KIB	2	R09-345	8	10/11/15

[a] Harvest date code
[b] Planting date code

Chapter 4

Evaluation of Pasteurization Methods to Preserve Edamame While Maintaining Acceptable Color and Texture

Abstract

Edamame is a food-grade soybean (*Glycine max* (L.) Merrill) that can be harvested either at the R6 (green with high moisture content) or R8 (mature and dry) reproductive growth stages. After edamame matures (R8), depending on the variety, the beans either retain their green color or turn yellow, black, or brown. If the green color remains after maturity, either the seed coat alone or the seed coat and cotyledon stay green. Edamame is a healthy alternative to snacks with high fat or sugar content; however, there is not a shelf-stable product commercially available. Pasteurizing in an acidic brine (pH <4.5) will require less thermal processing than sterilizing leading to a firmer texture with improved color. The objectives of this study were: i) improve shelf-stable edamame products at the R6 stage by evaluating pasteurization methods of high moisture edamame, and ii) evaluate the effect of pasteurizing colored edamame varieties in an acidic brine (pH <4.5) after harvesting at the R8 stage. For this research, three varieties were planted at the University of Arkansas Vegetable Research Station in Kibler, AR in 2016 and harvested at the R6 and R8 stages. A commercially available variety, 8080, was also used for Objective 1 with sugar and turmeric added as factors to retain the green color, and as an unprocessed commercial check for texture and color. The varieties harvested after maturity (R8) were also compared to their own unprocessed sample, commercially available canned black bean, pinto bean, and kidney beans. Texture was measured with a TMS 2000 texture analyzer, and color was measured with a HunterLab Color Flex. The color was interpreted as intensity of green color (IGC) and hue. The combination of pH and thermal processing in this research was sufficient to eliminate the threat *of C. boltulinum a*nd resulted in a commercially sterile product. The results of this research indicated green color is significantly ($p<0.05$) affected by turmeric, but not sucrose. In addition, the variety R07-10397 (green seed coat and cotyledon), harvested after maturing at the R8 stage, had the largest hue (most green) value, without the addition of sugar or turmeric. The varieties R07-589 (brown) and R09-345 (black) retained their color after processing and were similar in color to commercially available black,

pinto, and kidney beans. The texture of the pasteurized product was similar to the unprocessed commercial check. Pasteurizing green, black, or brown edamame beans after the mature growth stage (R8) may result in the most marketable, shelf-stable product.

Introduction

Edamame is a food-grade soybean (*Glycine max* (L.) Merrill) that is normally harvested by picking pods at the R6 reproductive stage: when the seeds are still green and fill 80-90% of the pods (Fehr et al., 1971; Konovsky et al., 1994; Shanmugasundaram and Yan, 2004). However, edamame can also be harvested when the seed is mature and dry at the R8 reproductive stage. While the seed color at R6 is green, the seed will either stay green or turn black, brown, or yellow at the R8 reproductive stage (Kiuchi et al., 1987). If the seed stay green at maturity, depending on the variety, either the seed coat alone or the seed coat and the cotyledon will remain green.

Two of the most important characteristics of edamame are color and texture. The standard for color is high green and low yellow content and the texture needs to be firm, but not chewy. (Funatsuki et al., 2006; IDA, 1990; Rackis, 1978; Watanabe, 1988).

Historically, edamame has been popular in countries such as Japan, China, Korea, and Taiwan (Shanmugasundaram and Yan, 2004); however, the edamame market is growing in the United States (Sharma, 2003; Zhang & Kyei-Boahen, 2007). Edamame is high in protein and phytochemicals and low in saturated fats. These healthy attributes make edamame a good food additive, such as an addition to soups and salads, or as healthy snack alternative to chips and candy (Masuda, 1991; Rayaprolu et al., 2015).

Edamame is preserved in several ways. Currently, the market primarily consists of either frozen, roasted (Mentreddy et al., 2002), or freeze-dried products (Rayaprolu et al., 2015). Frozen edamame, either in the pod or shelled, has high moisture content where roasted or freeze-dried have low moisture content. Prior to freezing, the product must be blanched to reduce enzyme activity (Mozzoni et al., 2009). Roasting or freeze-drying edamame provide

some advantages such as preserving a shelf stable product and retention of nutrients (Rayaprolu et al., 2015). Besides roasting and freeze-drying, pasteurizing edamame with high moisture content is another method to create a shelf stable product, hence, creating a new market for edamame.

The pigments causing the various colors at the R8 reproductive stage can attribute healthy benefits, such as preventing inflammation due to the anthocyanins and procyanidins in the black and brown seed coats (Kim et al., 2006; Nizamutdinova et al., 2009; Takahata et al., 2001). Thermal processing is required to preserve edamame with high moisture content. However, to produce a marketable product, the edamame must a have firm texture and acceptable color after thermal processing (Czaikoski et al., 2013; Mozzoni et al., 2009; Rayaprolu et al., 2015). Mozzoni et al. (2009) noted that increasing $CaCl_2$ will increase texture and increasing pH will decrease intensity of green color (IGC). The negative correlation between pH and IGC (or hue) is due to the acidity of the brine and heat process subtracting a Mg^2+ ion: converting the chlorophyll pigment into pheophytin (von Elbe & Schwartz, 1996). Reducing the thermal processing time can reduce this reaction resulting in a larger hue value and intensity of green color (Czaikoski et al., 2013). If the pH of the brine is less than 4.5, the edamame can be pasteurized instead of sterilizing requiring less thermal processing (Abbatemarco and Ramaswamy, 1994). Less thermal processing will result in a firmer texture (Czaikoski et al., 2013) and less break down of the chlorophyll (von Elbe & Schwartz, 1996). Czaikoski et al. (2013) estimated adding 3.43 g 100 mL^{-1} of sucrose can increase hue value. In addition, turmeric has been used in food preservation as a healthy alternative to synthetic food coloring (Abdeldaiem, 2014).

The first objective of this research was to improve shelf-stable edamame products at the R6 stage by evaluating pasteurization methods of high moisture edamame. Three varieties from the University of Arkansas were harvested and pasteurized at the R6 reproductive stage to determine differences in variety. In addition, a commercially available variety was preserved in a

brine consisting of different levels of turmeric and sugar to evaluate their effect on texture and color.

The second objective was to evaluate the effect of preserving edamame varieties that have either green, black, or brown seed coats at the R8 reproductive growth stage. The texture and color of the preserved edamame was compared to the product prior to processing as well as to preserved kidney, pinto, and black beans purchased from a local grocery store.

It was hypothesized that preserving green edamame (harvested at the R6 reproductive stage) by pasteurizing in an acidic brine will result in a product that has retained the green color and firm texture. It was also hypothesized that edamame varieties harvested at the R8 growth stage (mature and dry) with green, black, and brown seed coats can be preserved in acidic brine resulting in a product that is comparable to similar commercially available preserved beans.

Materials and Methods

Evaluation of Texture and Color

Texture

A single-bite test on a TMS 2000 texture analyzer (Food Technology Corp., Sterling VA, USA) with an Allo Kramer shear cell (10 blades) was used to measure texture. The settings of the instrument were max force at 50 kg, return distance at 40 mm, return speed at 3mm/sec, and contact force at 500g. Twenty grams of edamame were sampled for the sugar by turmeric experiment; however, 10 g were sampled for both experiments consisting of the three varieties R07-10397, R09-345, and R07-589 due to an increase in firmness of the processed edamame. The texture of the edamame was interpreted as the force in Newtons (N) the blades required to penetrate the sample with a single-bite test (Mozzoni et al., 2009).

Color

The color of the samples was measured with a HunterLab Color Flex (Hunter Associates Laboratory Inc., Reston, VA, U.S.A.). Three values were recorded, L*, a*, and b* which

represent the brightness/darkness, redness/greenness, and the blueness/yellowness of the sample, respectively. An increasing L* value indicates a brighter sample, a smaller a* value indicates a greener sample, and larger b* value indicates a more yellow sample (Hunter Associates Laboratory Inc., Reston, VA, U.S.A). The instrument was calibrated with a black glass tile first, then with a white standard tile. The white tile had L*, a*, and b* values of 93.76, -0.93, and 1.02, respectively. Prior to sampling, the calibration was validated with a green standard tile with values L* = 52.96, a* = -25.30, and b*=13.71. The intensity of green color and hue were calculated as (-a*/b*) (Mozzoni et al., 2009) and (degrees(ATAN2)(a*,b*)) (Rayaprolu et al. 2015), respectively. The intensity of green color indicates the ratio of green to yellow and the hue, measured in degrees, indicates how close the color is to pure red (0º), yellow (90º), green (180º), or blue (270º) as described by Lawless and Heymann (1998).

Base Brine

All edamame samples in the project were blanched in a 100º C water (tap) bath for 90 seconds to reduce 99% initial lipoxygenase activity (Mozzoni et al., 2009). The samples harvested at R6 were flash-frozen with liquid nitrogen to maintain cell structure during the freezing process (Luyet, 1968) after blanching. Prior to thermal processing, the samples were thawed by placing in an 82.2° C water bath and immediately cooled to ambient temperature by placing in cool water for 1 minute.

The base brine for both objectives consisted of a 0.95 L solution containing 0.53 L water, 0.38 L of 50 grain distilled white all-purpose vinegar, 57 grams of NaCl, and 2.5 grams of calcium chloride. The purpose of the vinegar was to bring the pH below 4.5 (Czaikoski et al., 2013) and the $CaCl_2$ was used to maintain a firm texture of the edamame after the thermal processing (Mozzoni et al., 2009).

Glass jars (236.59 mL) were purchased from the JarStore (www.JarStore.com). For each sample in this project, the jars were filled with 148.84 g of shelled and blanched edamame and 88.72 mL of brine.

For thermal processing, the closed jars were placed three-fourths into boiling water for 6 minutes. To ensure a commercially sterile pasteurized product, the jars were tested for pH and temperature. For pH, a pH Symphony SP79P (https://ca.vwr.com) instrument was used two weeks after processing. Two test jars were opened immediately after the thermal processing to check for a minimum temperature of 85° C (McGlynn, 2000) in the cold spot, located between one-third to one-half of the jar's height (Fellows, 2000; Mozzoni et al., 2009). After thermal processing, the jars were immediately cooled to ambient temperatures using tap water for 10 minutes.

Preservation of Edamame at the R6 Reproductive Stage

Variety Effect

Three varieties (fixed factor), R07-10397, R09-345, and R07-589, were planting in 2016 at the University of Arkansas Vegetable Research Station in Kibler, AR. The soil was a Dardanelle silt loam (fine-silty, mixed, superactive, thermic Typic Argiudolls) (Soil Survey, 2017). For the R6 harvest, entire plants were harvested and immediately transported to Fayetteville, AR and the pods were stripped from the plant using an edamame motive–power threshing machine (KE-6) (Doubletreasure Enterprise Inc.). The edamame was shelled from the pods using a "Little Sheller" (Taylor mfg. Co, Inc.). The edamame samples were placed in a refrigerator at 1.6º C and processed within 24 hours at the University of Arkansas' test kitchen.

Turmeric and Sugar Effect

The effect of sugar and turmeric on green color and texture was studied using a commercial variety '8080' from American Vegetable Soybean and Edamame, Inc. (AVS), an edamame company in Arkansas. Thermal processing was conducted at the Bryant Preserving Co. on August 31, 2016.

The effect of sugar and turmeric on color and texture was investigated with a two-factor factorial experimental design with three levels for both factors. The factors with fixed effects were sugar (0g 0.95 L^{-1}, 28g 0.95 L^{-1}, and 56 g 0.95 L^{-1}), oleoresin turmeric (emulsified solution)

(0 mL 0.95 L^{-1}, 0.25 mL 0.95 L^{-1}, 0.5 mL 0.95 L^{-1}), and the interaction of sugar and turmeric. The treatments were replicated three times (random effect). A frozen sample of 8080 was acquired from AVS and cooked in water at 100º C for six minutes. The color and texture of the unprocessed 8080 sample was used as a commercial check to the preserved edamame.

Preservation of Colored Edamame at R8 Reproductive Stage

Each treatment had three replications as a random effect. The varieties and seed colors were R07-10397 with green seed, R07-589 with brown seed, and R09-345 with black seed. All three varieties were planted at the University of Arkansas Vegetable Research Station in Kibler, AR in 2016. After harvesting at the R8 reproductive stage, the samples were stored in a cool, dry place in cloth bags until processing. Prior to processing, the R8 samples were soaked in distilled water for 24 hours to promote uniform texture and expansion during the thermal process (Nordstrom and Sistrunk, 1977). The processed product of the three colored varieties were compared to a pre-processed (soaked for 24 hours) sample and to a commercial product with similar color. The commercial sample of 8080 was used as a check to the R07-10397 (green) variety. The variety R07-589 was compared to unprocessed beans and to canned samples of pinto and kidney beans. The variety R09-345 was also compared to the beans before processing and to canned samples of black beans.

Statistical Analysis

Experimental factors were analyzed with SAS v.9.4 (SAS Institute, 2014) using the PROC MIXED procedure. Least square means (LSM) of the main effects and their interactions were estimated with the Type 3 method and the means were separated by interpreting the p values generated by the DIFF option.

Results

Commercial Sterility

After processing, the cold spot temperature and the pH of the brine was examined to ensure successful preservation. The cold spot temperature was above 85ºC and the average pH of the jars were 4.28.

Preservation of Edamame at the R6 Reproductive Stage

Variety Effect

Among the processed varieties, R07-10397 and R09-345 had the largest L* value ($p<0.05$) and R07-10397 had the lowest a* value, indicating greater green content. R07-10397 and R09-345 had a greater intensity of green color and hue ($p<0.05$) than R07-589. With a hue value of 87.57°, R07-10397 had a similar intensity of green color and hue as 8080 pasteurized with 0.25ml turmeric (88.56°) (Table 1). However, the frozen 8080 sample from AVS had a hue closer to true green than the preserved edamame (Table 1).

Turmeric and Sugar Effect

Adding sugar to the brine did not affect color or texture (Table 2); however, there were differences ($p<0.05$) on a*, b*, hue, and intensity of green color among the levels of turmeric (Table 3). Adding 0.5 mL 0.95 L^{-1} or 0.25 mL 0.95 L^{-1} of turmeric to the brine resulted in the lowest a* value, and the largest intensity of green color and hue value. The b* value was significantly smaller when adding turmeric at the level of 0.25 mL 0.95 L^{-1} versus 0.5 mL 0.95 $liters^{-1}$. When no turmeric was added, the edamame had a significantly lower intensity of green color and hue value ($p<0.05$) (Table 4). The processed edamame had a L* value similar to the frozen product; however, the frozen product had a hue value that was closer to a pure green color than the processed samples.

Preservation of Colored Edamame at R8 Reproductive Stage

R07-10397

The hue value (91.69°) and IGC (0.029) for the processed product of R07-10397 at R8 was lower, indicating less green color, than the frozen sample of 8080 and an unprocessed sample of R07-10397. However, the hue value of R07-10397, after processing, was similar to the processed edamame sample reported by Czaikoski et al. (2013) (93.50°) and was greater than 8080 (88.56°) after processing with sugar and turmeric (Table 5). The variety R07-10397 had the largest hue value without the addition of sugar or turmeric; therefore, this variety may give the best chance to preserve an edamame product with acceptable color. This variety matures with a green cotyledon in addition to a green seed coat, which may increase the seed's ability to retain a greater green color. Furthermore, the force value of R07-10397 did not differ ($p>0.05$) (408.94 N) compared to sample of 8080 (442.12 N).

R07-589 and R09-345

As the varieties R07-589 and R09-345 approach the R8 stage, R07-589 and R09-345 gradually change from green to brown and black, respectively. The L* value of R07-589 (17.26) prior to processing was lower ($p<0.05$) than the canned product of pinto (33.44) and kidney beans (28.51); however, after processing, the L* value of R07-589 (21.81) and the pinto and kidney beans did not differ. (Table 5). The hue value of the processed sample of R07-589 (44.38°) did not differ than that of the unprocessed sample and the canned samples of pinto and kidney beans (Table 5). There were no differences in L* value for R09-345 among all products (Table 5).

Discussion

Mozzoni et al. (2009) established a protocol to blanch edamame before sterilization in order to deactivate lipoxygenase activity, and developed a base brine consisting of NaCl and $CaCl_2$. Czaikoski et al. (2013) adapted the protocol set by Mozzoni et al. (2009) by pasteurizing in an acidic brine and evaluating levels of sucrose in an attempt to retain green color. Czaikoski

et al. (2013) concluded that the beans processed with sucrose were significantly greener than without; however, the processed product was significantly less green than the beans *in natura*. Although McGlynn et al. (1993) reported a brine consisting of a pH below 4.5 can result in a firmer texture after thermal processing, Czaikoski et al. (2013) observed a product that was less firm than beans *in natura,* when pasteurizing in an acidic brine.

The overall objective of this research was to improve the methodologies established by Mozzoni et al. (2009) and Czaikoski et al. (2013) resulting in a product that would be commercially acceptable. The combination of pH and thermal processing in this research was sufficient to eliminate the threat of *C. boltulinum* and resulted in a commercially sterile product (McGlynn, 2000). Although the brine was under the acidic (pH of 4.5) threshold, adding a greater concentration of vinegar to bring the pH below 4.0 would give more assurance of a safe product (McGlynn, 2000). Czaikoski et al. (2013) estimated adding 3.43 g 100 mL^{-1} of sucrose can significantly ($p<0.05$) increase the hue value by 1.17º. Although this research did not find sucrose to have a significant effect on color, the hue value actually decreased (less green) by 0.34º after adding 56g 0.95 L^{-1} of sugar. The discrepancy between the two results may be due to the fact the suggestion of 3.43 g 100 mL^{-1} reported by Czaikoski et al. (2013) was a projection as 3.43 g 100 mL^{-1} was outside of the central composite design to evaluate the effects of added sucrose.

Abdeldaim (2014) suggested adding turmeric can be a healthy alternative to artificial dyes, as turmeric can have health benefits, such as antioxidant and antimicrobial activities. Furthermore, Cleary and McFeeters (2006) inferred using turmeric in a pasteurization process for dill pickles can minimize off-flavors due to oxidation. The results of this project indicate adding 0.25 mL 0.95 $L^{-1}$of turmeric would result in a hue closer to green compared to no turmeric. When comparing the two pasteurization processes using edamame harvested at the R6 stage, it was discovered the variety R07-10397 had a similar hue value (87.57), without the addition of turmeric, compared to the variety 8080 with 0.25 mL 0.95 L $^{-1}$of turmeric added (88.56).

The variety R07-10397 (green seed coat and cotyledon), harvested after maturing at the R8 growth stage, had the largest hue (most green) value without the addition of sugar or turmeric; therefore, R07-10397 may give the best chance to preserve an edamame product with acceptable color. This variety matures with a green cotyledon in addition to a green seed coat, which may increase the seed's ability to retain a higher green color.

The brightness (L*) value of R07-589 harvested mature and dry (R8 growth stage) improved after processing. The increase in brightness after processing agrees with the results noted by Rayaprolu et al. (2015).

Furthermore, after processing, the hue value of R07-589 was similar to the unprocessed sample, canned pinto beans, and canned kidney beans. These results indicate thermal processing in an acidic brine did not significantly alter the red-brown color of R07-589; furthermore, R07-589 had a color statistically similar to canned pinto and kidney beans.

Although the processed sample of R09-345 had a lower L* value (darker) than the unprocessed sample and the canned black beans, there were no differences ($p>0.05$). This indicates the thermal process in an acidic brine did not alter the black color of the beans.

The retention of red-brown and black color for R07-589 and R09-345, respectively, after processing indicates the pigments causing the colors did not break down due to the thermal process or the acidic brine leaving a product that would be aesthetically acceptable with healthy attributes.

Lau et al. (2000) reported vegetables will soften during thermal processing. In this study, however, the texture of the processed edamame in the sugar by turmeric test (R6 growth stage) and the three mature (R8) varieties were similar to the commercial check (8080). Maintaining the texture can be attributed to the addition of $CaCl_2$ (Mozzoni et al., 2009) and decrease in duration of thermal processing (Czaikoski et al., 2013).

Conclusions

The main objective of this research was to find a method to preserve edamame with high water content to be shelf-stable at room temperature and retain acceptable color and texture. The variety R07-10397 harvested at the R8 reproductive stage, without additives for color, resulted in a value closest to green and had a texture similar to the frozen product of 8080. The mature samples of the varieties R07-589 and R09-345 retained their brown and black color after processing and were similar in color and texture to their respective commercial check. The results of this research suggest preserving edamame after the R8 (mature and dry) growth stage can have commercially acceptable color (green, brown, and black) and texture, and adding turmeric can help maintain green color. The pigments from the different colors and the addition of turmeric increase the viability of these products by adding health benefits.

References

Abbatemarco, C., and H.S. Ramaswamy. 1994. End-over-end thermal processing of canned vegetables: effect on texture and color. Food Res. Int. 27:327-334.

Abdeldaiem, M.H. 2014. Use of yellow pigment extracted from turmeric (*Curcuma Longa*) rhizomes powder as natural food preservative. Am. J. Food Sci. Technol. 2:36-47.

Czaikoski, K., R.S. Leite, J.M.G. Mandarino, M.C. Carrao-Panizzi, J.B. da Silva, and E.I. Ida. 2013. Canning of vegetable-type soybean in acidified brine: effect of the addition of sucrose and pasteurization time on color and other characteristics. Ind. Crops Products. 45:472-476.

Cleary, K., and R.F. McFeeters. 2006. Effects of oxygen and turmeric on the formation of oxidative aldehydes in fresh-pack dill pickles. J. Agric. Food Chem. 54:3421-3427

von Elbe, J.H., and S.J. Schwartz. 1996. Colorants. In: Food Chemistry, 3rd edn (edited by O. Fennema). pp. 651-722, New York, USA: 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.) Merrill. Crop. Sci. 11:929-931

Funatsuki, H., M. Ishimoto, H. Tsuji, K. Kawaguchi, M. Hajika, and K. Fujino. 2006. Simple sequence repeat markers linked to a major QTL controlling pod shattering in soybean. Plant Breed. 125:195-197.

Fellows, P. 2000. Heat Sterilization. In: Food processing technology: principles and practice, 2nd edn. (edited by P. Fellows). pp. 250-277, Boca Raton, USA: CRC Press LLC.

IDA. 1990. Iwate-ken Seikabutsu To Shukka Kikaku Shido Hikkei. Iwate-ken Department of Agriculture, Morioka.

Kiuchi, Y., H. Ishikawa, N. Nitta, and T. Sasaki. 1987. Collection of crop genetic resources and their characteristics in Iwate Prefecture In: characteristics of local soybean varieties. *Tohoku Nogyo Kenkyu* 40:139-40.

Kim, E.H., S.H. Kim, J.I. Chung, H.Y. Chi, J.A. Kim, and I.M. Chung. 2006. Analysis of phenolic compounds and isoflavones in soybean seeds (Glycine max (L.) Merrill) and sprouts grown under different conditions. Eur. Food Res. Technol. 222:201 https://doi.org/10.1007/s00217-005-0153-4.

Konovsky, J., T.A. Lumpkin, and D. McClary. 1994. Edamame: the vegetable soybean. pp. 173-81 in A.D. O'Rourke (ed), Understanding the Japanese food and agrimarket: a multifaceted opportunity. Haworth Press, Binghamton.

Lau, M.H., J. Tang, and B.G. Swanson. 2000. Kinetics of textural and color changes in green asparagus during thermal treatments. J. Food Engineering 45:231-236.

Lawless, H.T., and H. Heymann. 1998. Sensory evaluation of food. Chapman and Hall, New York. p. 819.

Luyet, B.J. 1968. Investigations on freezing and freeze-drying of selected fruits and vegetables. Technical report 69-8-FL. U.S. Army Natick Laboratories, Natick, MA.

Masuda, R. 1991. Quality requirement and improvement of vegetable-type soybean. In: Shanmugasundaram, S. (Ed.), Vegetable-type soybean: Research Needs for Production and Quality Improvement. Asian Vegetable Research and Development Center, Taiwan. 92-102.

McGlynn, W. 2000. 'The Importance of food pH in commercial canning operations". Division of Agricultural Sciences and Natural Resources. Oklahoma State University Cooperative Extension Service, Web. 14 November 2017. < http://pods.dasnr. okstate.edu/ docushare/dsweb/Get/Document-603/FAPC-118pod.pdf>.

McGlynn, W.G., D.R. Davis, and F., Honarmand. 1993. Gluconic acid influences texture and color of canned asparagus. J. Food Sci. 58:614-615.

Mentreddy, S.R., A.I. Mohamed, N. Joshee, and A.K. Yadav. 2002. Edamame: A nutritious vegetable crop. pp. 432-438. In: J. Janick and A. Whipkey (eds.), Trends in new crops and new uses. ASHS Press, Alexandria, VA.

Mozzoni, L.A., R.O. Morawicki, and P. Chen. 2009. Canning of vegetable soybean: procedures and quality evaluations. International J. Food Sci. Technol. 44:1125-1130.

Nizamutdinova, I.T., Y.M. Kim, J.I. Chung, S.C. Shin, Y.K. Jeong, H.G. Seo, et al. 2009. Anthocyanins from black soybean seed coats stimulate wound healing in fibroblasts and keratinocytes and prevent inflammation in endothelial cells. Food Chem. Toxicology 47:2806-2812.

Nordstrom, C.L., and W.A. Sistrunk. 1977. Effect of type of bean, soak time, canning media and storage time on quality attributes and nutritional value of canned dry beans. J. Food Sci. 42:795-798.

Rackis, J.J. 1978. Biochemical changes in soybeans: maturation, post-harvest storage and processing, and germination. pp. 34-76. *In*: H.O. Hultin and M. Milner *(eds*), Post-harvest Biology and Technology. Food and Nutrition, Westport.

Rayaprolu, S., N. Hettiarachchy, M. Aldoury, S. Cho, D. Moseley, and P. Chen. 2015. Physical and textural attributes of freeze-dried genetically modified and non-genetically modified soy beans. J. Food Nutr. Sci. 3:119-125.

SAS Institute. 2014. The SAS System for Windows. Version 9.4. SAS Inst. Inc., Cary, NC.

Shanmugasundaram, S., and Yan MR. 2004. Global expansion of high value vegetable soybean. VII world soybean conference/IV international soybean processing and utilization conference/III Brazilian Soybean Congress, 29 Feb–5 Mar 2004. *Foz do Iguassu,* Brazil, pp 915–920.

Soil Survey Staff, Natural Resources Conservation Service, United States Department of Agriculture. Web soil survey. Available online at the following link: https://websoilsurvey.sc.egov.usda.gov/. Accessed [12/11/2017].

Sharma, K.P., 2003. Varietal yield stability and appropriate management for quality organic edamame production. Project report. Organic sector development program (1-166.) The sharing farm Society – Richmond Food Security Society. Nature's path foods. Retrieved from http://certifiedorganic.bc.ca/programs/osdp/1-166_Edammae2_Final_Report.pdf.

Takahata, Y., M. Ohnishi-Kameyama, S. Furuta, M. Takahashi, and A. Suda. 2001. Highly polymerized procyanidins in brown soybean seed coat with a high radical-scavenging activity. J. Agric. Food Chem. 49:5843-5847.

Watanabe, M. 1988. Tunneru Saibai – Sapporo-midori. Nogyo Gijutsu Taikei (yasai) 13:1-8.

Zhang, L., and S. Kyei-Boahen. 2007. Growth and yield of vegetable soybean (edamame) in Mississippi. Hortic. Technol. J. 17:26-31.

Table 1. Color values of three edamame varieties harvested at the R6 reproductive stage after pasteurization, and an unprocessed sample of a commercial variety as a check.

Effect	L*	a*	b*	IGC[f]	Hue(º)	Force (N)
Pickled						
R07-10397(R6)	46.75[a]	1.26[b]	38.72[c]	-0.04[b]	87.57[b]	310.08[c]
R07-589(R6)	34.27[b]	9.37[d]	19.86[a]	-0.47[c]	64.78[c]	355.11[b]
R09-345(R6)	45.94[a]	4.78[c]	23.95[ab]	-0.20[b]	78.70[b]	345.70[b]
Check						
8080[g]	47.86[a]	-10.09[a]	26.04[b]	0.39[a]	111.07[a]	446.76[a]

Means followed by the same letters are not significantly different (alpha = 0.05) according to Fisher's protected least significant differences (LSD).

[f] Intensity of green color (-a/b)

[g] 8080 average values from samples from AVS in 2015 and 2016

Table 2. Levels of sucrose added to the brine, and color values after pasteurization.

Effect	Amount	Hue°	Force (N)
Sugar	0g[a]	88.68[a*]	470.4[a]
Sugar	28g[a]	88.41[a]	454.4[a]
Sugar	56g[a]	88.34[a]	465.7[a]

*Means followed by the same letters are not significantly different (at 5% probability) according to Fisher's protected least significant differences (LSD).

[a] The concentration of sugar was measured as g 0.95 L^{-1}

Table 3. Analysis of Variance (ANOVA) table displaying the mean square (MS), residual mean square (RMS), F values, and significance level for color values for the turmeric factor.

Source	MS	RMS	F Value	Pr > F
Turmeric (a*)	2.629	0.105	25.08	<0.0001
Turmeric (b*)	490.069	3.667	133.66	<0.0001
Turmeric (Hue)	9.049	0.139	64.68	<0.0001
Turmeric (IntGreenColor)	0.003	4E-05	64.80	<0.0001

No values are significant (5% probability) for L* and Force (N)
All values of Turmeric have DF = 2

Table 4. Color values for the variety 8080 after pasteurization in an acidic brine with different levels of the turmeric factor, and an unprocessed sample of a commercial variety as a check.

Turmeric	a*	b*	IGC	Hue^{o}	L*	Force (N)
Pickled						
0 mL[b]	1.96c*	37.57b*	-0.052c*	87.01c*	51.25a*	472.28a*
0.25 mL	1.15b	47.02c	-0.025b	88.56b	51.01a	452.46a
0.5 mL	0.944b	51.36d	-0.020b	88.88b	51.52a	465.77a
Check						
8080[a]	-10.62a	26.38a	0.40a	111.95a	42.94a	436.03a

*Values with the same letter are not significantly different ($p>0.05$)
L* and Force (N) are not sig different among levels of turmeric
[a]8080 average values from samples from AVS in 2015 and 2016
[b] The concentration of turmeric was measured as mL 0.95 L^{-1}

Table 5. Color values of three colored varieties harvested at the mature and dry (R8) reproductive growth stage. The varieties and colors are: R07-10397 (green) R07-589 (brown) and R09-345 (black).

Effect	L*	a*	b*	IGC[d]	Hue[o]	Force (N)
R07-10397						
R07-10397	49.08[a]	-0.6913[c]	30.2995[b]	0.029[c]	91.69[c]	408.94[a]
R07-10397 DS[e]	46.18[a]	-7.005[b]	29.1228[ab]	0.249[b]	103.96[b]	.
8080	49.54[a]	-10.158[a]	26.9195[a]	0.382[a]	110.92[a]	442.12[a]
R07-589						
R07-589	21.81[ab]	19.16[c]	18.61[ab]	-1.022[ab]	44.38[ab]	461.73[a]
R07-589 DS[e]	17.26[b]	13.67[a]	13.08[a]	-1.022[ab]	44.44[ab]	.
PINTO	33.44[a]	17.21[b]	21.83[b]	-0.829[b]	50.39[b]	117.01[b]
KIDNEY	28.51[a]	20.19[c]	18.83[b]	-1.114[a]	41.88[a]	112.15[b]
8080	.	.	.	.	.	435.37[a]
R09-345						
R09-345	8.9[a]	6.52[b]	-0.19[a]	48.963[a]	2.07[ab]	476.75[a]
R09-345 DS[e]	10.3[a]	0.30[a]	-0.45[a]	-36.50[b]	-42.92[b]	.
Black	13.4[a]	9.53[c]	5.45[b]	-1.873[ab]	31.08[a]	77.81[c]
8080	.	.	.	.	.	421.63[b]

Means followed by the same letters are not significantly different (at 5% probability) according to Fisher's protected least significant differences (LSD).
[d] Intensity of green color (-a/b)
[e] Dry soaked beans

Overall Conclusion

In conclusion, the results of this dissertation will contribute to defining breeding, agronomic, and processing strategies to increase domestic production of edamame in the United States. The edamame market is growing in the U.S.; however, a majority of the edamame is imported from other countries. More-adapted varieties are necessary to increase domestic production. The evaluation of diversity among accessions from seven countries and discovering QTL association with seed weight and size traits will enable breeders to develop larger and more-adapted varieties. Furthermore, the observations of harvest windows for edamame will assist private companies and famers to manage their crop more efficiently. Finally, an improved preserving technique will result in a higher quality, shelf-stable product.

In 2012, a private edamame company began what is thought to be the first edamame production and processing company in the United States. The location they chose to build a processing plant was Mulberry, Arkansas. They began production with an edamame variety from China and a variety released from the University of Arkansas. It has been a challenging endeavor to develop a variety that has improved yield and size. It is our hope that the molecular markers observed in this research will help breed and select for improved varieties. There have been other private companies inquire into our edamame varieties since 2012, giving evidence that demand for improved varieties will remain high.

It is our understanding that predicting the most optimum harvest for edamame at the R6 reproductive stage remains more of a guessing game than science. The results of this research begins to record data to help predict the best harvest time of edamame, and gives an approximate window of how long an edamame field will remain green after the pods reach the largest collective weight. Caution should be taken as it is understood (for some varieties), the beans can begin to turn yellow before the pods. Future research will include drones equipped with cameras that can view the pods at R6 through the canopy. This will allow the researcher to

collect data on a daily basis. As data is collective across more years, locations, and maturity groups, we will be able to schedule a harvest with more certainty.

Recently, there have been an increase of shelf-stable edamame products developed; however, these products are either roasted or freeze-dried. A need remains for edamame, with a high moisture content, to be shelf-stable. Shelf-stable high-moisture edamame can be used as a topping for foods such as soups and salads. The results of this research show the variety R09-345 (black seed coat) retains the black color after preserving, and may be a viable product to develop.