QTL MAPPING AND OPTIMUM RESOURCE ALLOCATION FOR ENHANCING AMINO ACID CONTENT IN SOYBEAN

by

CALEB VAN WARRINGTON IV

(Under the Direction of H. Roger Boerma)

ABSTRACT

Soybean [*Glycine max* (L.) Merr.] is the world's leading oilseed crop and provider of high quality protein meal. The protein portion of the soybean seed represents an energy-efficient source of amino acids for use in animal diets. As a sole dietary source of protein for poultry and swine, soybean is deficient in the sulfur-containing amino acids methionine (Met), cysteine (Cys), and may also have sub-optimum levels of threonine (Thr), and lysine (Lys). Enhancing these essential amino acids would improve the nutritive value of soybean meal and provide additional value to the animal feed industry. In this study, a population of recombinant inbred lines (RILs) from the cross of 'Benning' × 'Danbaekkong' was used to investigate the inheritance of quantitative trait loci (QTL) associated with protein and amino acid concentrations and the presence and magnitude of genotype × environment interaction for these traits. The RILs were grown in five field environments. QTL were detected for crude protein (cp), Lys/cp, Thr/cp, Met/cp, Cys/cp, and Met+Cys/cp using DNA markers. The Danbaekkong allele at a major protein QTL was found to be associated with reduced levels of each of the amino acids. Selection for amino acid QTL on other chromosomes may increase protein quality and retain a high level of overall crude protein. The effect of genotype × environment was relatively minor

for each trait based on an assessment of the variance components. The estimated the number of environments and replications necessary to detect certain differences between two genotype means was determined. Results indicated that five replications and two environments could detect a difference of 2.5% between two RIL means for Lys and Thr. An increased number of plots (environment/replication combinations) would be necessary to detect a 2.5% difference or less between two RIL means for crude protein, Met/cp, Cys/cp, and Met+Cys/cp. This information would be useful in developing the most cost-effective and efficient testing scheme for these traits in a breeding program for these traits. This research should increase the understanding of the genetic basis for protein and specific amino acids and provide for effective and efficient genetic improvements for these traits.

INDEX WORDS: soybean, lysine, threonine, methionine, cysteine, amino acids, recombinant inbred lines, quantitative trait loci, genotype × environment interaction

QTL MAPPING AND OPTIMUM RESOURCE ALLOCATION FOR ENHANCING AMINO ACID CONTENT IN SOYBEAN

by

CALEB VAN WARRINGTON IV

BS, University of Delaware, 2001

MS, University of Georgia, 2006

A Dissertation Submitted to the Graduate Faculty of The University of Georgia in Partial Fulfillment of the Requirements for the Degree

DOCTOR OF PHILOSOPHY

ATHENS, GEORGIA

2011

QTL MAPPING AND OPTIMUM RESOURCE ALLOCATION FOR ENHANCING AMINO ACID CONTENT IN SOYBEAN

by

CALEB VAN WARRINGTON IV

Major Professor:	H. Roger Boerma
Committee:	E. Charles Brummer
	John N. All
	Paul L. Raymer
	David A. Knauft

Electronic Version Approved:

Maureen Grasso
Dean of the Graduate School
The University of Georgia
May 2011

DEDICATION

This dissertation is dedicated to my family for all of their support.

ACKNOWLEDGEMENTS

These acknowledgements are in no particular order (other than acknowledging my committee first, for obvious reasons): Roger Boerma, John All, Charlie Brummer, Paul Raymer, David Knauft, Dean Kemp, Kurk Lance, Dale Wood, Herschel Chambers, Erin Leggett, Jennie Alvernaz, Bo Keun Ha, my parents, Chris Beaumont, Samantha Shelton, Hussein Abdel-Haleem, Wayne Parrott, Donna Harris, Earl Baxter, Gina Rowan, David Walker, David Hulburt, Shuquan Zhu, Chris Collins, Zack Shearin, Aaron Hoskins, Jennifer Yates, Maria Monteros, and Jim Turner.

TABLE OF CONTENTS

LIST OF TABLES

Page

LIST OF FIGURES

Page

CHAPTER 1

INTRODUCTION

Soybean [(*Glycine max* (L.) Merr.] is a legume species native to East Asia in the Fabaceae family. It is a subtropical plant but is grown as far north as southern Canada in North America. The genus *Glycine* has two subgenera, *Soja* (Moench) and *Glycine*. The subgenus *Soja* (Moench) includes the cultivated soybean (*G. max*) and the wild soybean *G. soja* (L.) Sieb and Zucc). Both are annual plants with identical chromosomal arrangements ($2n = 2x = 40$) and are cross-fertile (Hymowitz 1970). *Glycine soja* is most likely the progenitor of cultivated soybean and still grows wild in China, Japan, Korea, Taiwan, and Russia. Historical analyses have shown that domestication likely occurred during the Chou Dynasty, around the 11th century B.C. in northeast China (Hymowitz, 1970).

The efforts of breeders and geneticists in the early part of the 20th century allowed for soybean to emerge as one of the most important agricultural crops worldwide. Once grown in North America primarily as a forage crop, soybean is now the world's leading source of edible protein, vegetable oil, phospholipids, dietary antioxidants (such as tocopherols and isoflavones), nutraceuticals (such as sterols), and other ingredients of foods, feeds, and industrial products (Bellis, 2003). In the recent past, soybean production has expanded from North America into South America and China as the demand for soybean and soybean products has increased. Brazilian soybean production is expected to increase 25-50% by 2012 based on 2003 estimates (Durham, 2003). This expansion will lead to a very competitive global market and even greater demand for the continued improvement of soybean through breeding and biotechnological tools.

Soybean prices were at their highest in nearly 20 years during the last few years as more emphasis is being placed the utilization of plant biproducts as alternative sources of fuel (eg. soy biodiesel). In addition, soy foods have become more popular in our more health-conscious society. Studies have shown that the incorporation of soybean into our daily diets may aid in the prevention of disease, including coronary heart disease, and may serve as a cholesterol reducing agent (Hasler 2002).

Soybean breeding involves the selection of parental material, the creation of a segregating population, advancing the population towards homozygosity (either with or without selection), evaluating relatively homozygous lines, and finally the release of a pure-line cultivar. Selection can occur in early generations or later among inbred lines. The effectiveness of selection depends on heritability of the trait and environment(s) where the populations or lines are grown. Breeding populations are usually developed by hybridization of two or more parents (cultivars, breeding lines, or other germplasm). Populations are advanced through generations of inbreeding by self pollination. Nearly homozygous lines are then created from individually harvested inbred plants. There are several procedures utilized by soybean breeders as they advance crosses to homozygosity (Orf et al., 2004). These include pedigree selection, single seed descent, and the bulk method. Pedigree selection is based on the selection among and within the best appearing families in each generation. Single seed descent involves advancing a single seed or pod from each plant to the next generation (Brim, 1966). This method is useful in that it allows for the development of nearly homozygous lines quickly while still preserving most of the original genetic variation within the population. In the bulk method, populations are advanced in bulk without artificial selection until later generations, at which time nearly homozygous lines are selected for seed yield testing. Other procedures used in soybean cultivar

development include early generation testing, in which testing for yield starts in the F_2 or F_3 generation and backcrossing, where a donor parent is continually backcrossed to a recurrent parent in an effort to introgress a chromosomal segment. Recurrent selections aims to improve the population over multiple cycles of selection (Bernardo, 2003; Orf et al., 2004; St Martin and Geraldi, 2002).

The USDA maintains roughly 15,000 accessions of *Glycine max* and a few other *Glycine* species in Urbana, Illinois, and Stoneville, Mississippi (Carter Jr et al., 2004). Brazil, China, Germany, India, Indonesia, Japan, Russia, and South Korea have germplasm collections as well, with the total *G. max* accessions estimated at over 100,000. Germplasm resources are of great importance in soybean breeding efforts as there is a narrow genetic base in modern soybean cultivars grown in the USA (Hyten et al., 2006). It has been reported that most of the soybean cultivars grown in the USA are the ancestors of less than 15 introduced lines (Gizlice et al., 1993). Studies have also shown that there is significant genetic variation between the northern and southern soybean gene pool in the USA (Li et al., 2001; Sneller, 1994). Sneller (1994) looked at 122 northern and southern U.S. lines and found that the southern lines derive 73.2% of their parentage from only six ancestors, and the lines in the northern USA derive 59.7% of their parentage from six ancestors. The introgression of alleles from wild germplasm that confer resistance to diseases or pests has been undertaken with much success by soybean breeders. Plant introductions (PIs) were evaluated for root-knot nematode resistance and lines were then used to transfer resistance alleles to elite germplasm (Luzzi et al., 1987; Tamulonis et al., 1997). Three Japanese PIs, PI 171451 (‘Kosamame’), PI 229358 (‘Soden-daizu’), and PI 227687 (‘Miyako White’), were found to confer resistance to many soybean insect pests (Lambert, 1984; Luedders, 1977) therefore, they have been utilized as the donors of resistance alleles in numerous

breeding programs where insect damage is costly to growers (Boethel, 1999; Lambert and Tyler, 1999). A wild soybean accession, *G. soja*, also proved useful in the discovery and utilization of a major quantitative trait locus (QTL) for high protein (Chung et al., 2003). Harris (2001) also reported this allele to be present in the South Korean cultivar 'Danbaekkong', which has been utilized in the soybean breeding program at the University of Georgia.

The major breeding objectives for soybean include seed yield, agronomic traits i.e. lodging and plant height), adaptation traits (i.e. maturity), disease and insect resistance, and seed compositional and quality traits. The focus of this research concerns protein quality in soybean. As soybean is extensively used as a high-protein feed in livestock and poultry production, its amino acid composition is critical to ensuring optimal growth and development of animals. As a sole source of protein as well as in combination with grain rations, soybean is deficient in the sulfur-containing amino acids methionine (Met) and cysteine (Cys) as well as threonine (Thr) and lysine (Lys). Increasing these amino acids through biotechnological tools such as transformation has been met with limited success (Altenbach et al., 1987; Kortt et al., 1991). Therefore, plant breeding approaches are being utilized in order to improve concentrations of these amino acids in soybean.

The objective of our first study was to identify genomic regions associated with crude protein and amino acid concentrations using the association of these traits with simple sequence repeat (SSR) and single nucleotide polymorphism (SNP) markers in a population derived from a cross between a high-protein cultivar 'Danbaekkong' and a cultivar with average protein content 'Benning'. The identification of markers linked to quantitative trait loci (QTL) conditioning enhanced levels of protein and amino acids would provide breeders with the tools to implement a marker-assisted selection (MAS) breeding strategy.

The second objective of this research was to evaluate the nature of genotype × environment interaction for protein and amino acid concentration in a soybean population. Understanding G×E interactions is an integral part of efforts to breed for soybean with improved protein quality traits in soybean. The nature of this interaction, or lack thereof, can influence how breeding efforts proceed. If no significant interactions exist, it would be feasible to test genotypes at one location during one year. G×E interaction greatly impacts breeders' ability to successfully select for traits of interest, as genotypic means are dependent upon the environment in which they are grown. In addition, the number of growing environments necessary to provide adequate selection precision greatly impacts the allocation of time and resources for testing across multiple years.

References

Altenbach S.B., Pearson K.W., Leung F.W., Sun S.S.M. (1987) Cloning and sequence analysis of a cDNA encoding a Brazil nut protein exceptionally rich in methionine. Plant molecular biology 8:239-250.

Bellis D., H.R. Boerma, E. Ready, and R. Wilson. (2003) Strategic Plan for Soybean Genomics, Proceedings of Soybean Genomics Research Workshop II, St Louis.

Bernardo R. (2003) On the effectiveness of early generation selection in self-pollinated crops. Crop Sci. 43:1558-1560. DOI: 10.2135/cropsci2003.1558.

Boethel D. (1999) Assessment of soybean germplasm for multiple insect resistance. Global plant genetic resources for insect-resistant crops. CRC, Boca Raton, FL:101–129.

Brim C. (1966) A modified pedigree system of selection in soybeans. Crop Sci 6:220.

Carter Jr T., Nelson R., Sneller C., Cui Z. (2004) Genetic diversity in soybean. Soybeans: Improvement, Production, and Uses, Ed 3:303–416.

Chung J., Babka H., Graef G., Staswick P., Lee D., Cregan P., Shoemaker R., Specht J. (2003) The seed protein, oil, and yield QTL on soybean linkage group I. Crop Science 43:1053.

Durham D. (2003) The United Soybean Board's better bean initiative: Building United States soybean competitiveness from the inside out.

Gizlice Z., Carter T.E., Burton J.W. (1993) Genetic Diversity in North American Soybean: I. Multivariate Analysis of Founding Stock and Relation to Coefficient of Parentage. Crop Sci. 33:614-620. DOI: 10.2135/cropsci1993.0011183X003300030038x.

Harris D.K. (2001) Genetic improvement of soybean seed traits and resistance to bud blight and root-knot nematodes, Crop and Soil Sciences, University of Georgia, Athens, GA.

Hyten D.L., Song Q., Zhu Y., Choi I.Y., Nelson R.L., Costa J.M., Specht J.E., Shoemaker R.C., Cregan P.B. (2006) Impacts of genetic bottlenecks on soybean genome diversity. Proceedings of the National Academy of Sciences 103:16666.

Kortt A., Caldwell J., Lilley G., Higgins T. (1991) Amino acid and cDNA sequences of a methionine-rich 2S protein from sunflower seed (Helianthus annuus L.). European journal of biochemistry/FEBS 195:329.

Lambert L., Tyler J. (1999) Appraisal of insect-resistant soybeans. Economic, environmental, and social bene ts of insect resistance in eld crops. Thomas Say Publication, Entomological Society of America, Lanham, MD.

Lambert L.K. (1984) Multiple Insect Resistance in Several Soybean Genotypes1. Crop Science 24:887.

Li Z., Qiu L., Thompson J.A., Welsh M.M., Nelson R.L. (2001) Molecular Genetic Analysis of U.S. and Chinese Soybean Ancestral Lines. Crop Sci. 41:1330-1336. DOI: 10.2135/cropsci2001.4141330x.

Luedders V. (1977) Resistance of selected Soybean Genotypes and Segregating Populations to Cabbage Looper Feeding1. Crop Science 17:395.

Luzzi B., Boerman H., Hussey R. (1987) Resistance to three species of root-knot nematode in soybean. Crop science (USA).

Orf J., Diers B., Boerma H. (2004) Genetic improvement: Conventional and molecular-based strategies. Soybeans: Improvement, production, and uses. 3rd ed. Agron. Monogr 16:417–450.

Sneller C.H. (1994) Pedigree Analysis of Elite Soybean Lines. Crop Sci. 34:1515-1522. DOI: 10.2135/cropsci1994.0011183X003400060019x.

St Martin S.K., Geraldi I.O. (2002) Comparison of three procedures for early generation testing of soybean. Crop Science 42:705-709.

Tamulonis J., Luzzi B., Hussey R., Parrott W., Boerma H. (1997) DNA markers associated with resistance to Javanese root-knot nematode in soybean. Crop Science 37:783-788.

CHAPTER 2

LITERATURE REVIEW

Soybean Protein and Amino Acids

As the world population increases, so too will the demand for animal protein. It is projected that global meat production and consumption will increase from 233 million tonnes in 2000 to 300 million t by the year 2020 (Annicchiarico, 2002). Soybean [*Glycine max* (L.) Merr.] is the world's leading oilseed crop and provider of high-quality protein meal. Soybean meal, a by-product of oil extraction, drives the soybean market due mainly to its use as a source of amino acids in livestock and poultry rations, an industry which uses roughly 77% of soybean meal (Kerley and Allee, 2003). With this in mind its importance to the U.S. agriculture industry cannot be overstated. Soybean meal is a concentrated source of protein and energy and has lower crude fiber content than most other oilseed meals, allowing nutritionists to formulate higher energy diets which are more efficient in the conversion of feed to meat (Smith, 1997). Although soybean protein is the most prominent among the oilseed crops, it is far from a source of complete protein nutrition and thus bears improvement. Improving the composition of soybean seed would add economic value along the entire value chain, from grower to end-user.

Protein quality refers to the balance and composition of amino acid constituents which comprise overall protein. Protein, *per* se, although important is not as critical as its quality. The major function of protein in nutrition is to supply adequate amounts of required amino acids (Friedman and Brandon, 2001). As a sole dietary source of protein, soybean is deficient in the amino acids methionine (Met), cysteine (Cys), threonine (Thr), and lysine (Lys). Each of these

amino acids is considered essential, as monogastric animals (ie. swine and poultry) cannot synthesize them *de novo* and therefore each must be obtained solely from the diet. Any deficiency in the amino acid balance must be supplemented in the diet at additional costs to livestock producers, a process which costs approximately $100 million annually (Imsande, 2001). Clarke and Wiseman (2000) speculated that a 10% increase in Lys, Met, and Thr concentrations would yield a $4.5 to 9.5, $2.7, and $5.9/T increase in commercial meal value, respectively. Besides the associated costs Met supplementation may cause additional problems such as leaching during soybean meal processing and bacterial degradation leading to the formation of undesirable volatile sulfides (George and De Lumen, 1991). Therefore, the development of soybean cultivars with enhanced amino acid balance would increase their economic value and reduce any negative environmental effects associated with supplementation. The United Soybean Board (USB) consists of 69 farmer-directors from 29 soybean-producing states that oversee the investments of the National Soybean Checkoff program. The checkoff program first collects and then invests the funds to advance soybean marketing, production technology, and the development of new uses (soybeancheckoff.com). The United Soybean Board's Better Bean Initiative (BBI) included as one of its major research goals in 2002 that soybean breeders strive towards developing soybean cultivars in the USA with improved oil and meal traits so as to better compete with foreign producers and provide customers with desirable products (Durham, 2003). One of the primary traits targeted by the BBI is increased Met+Cys concentration, while increased levels of Lys and Thr represent secondary goals of the initiative.

The value of increased levels of these amino acids is application driven, as swine and poultry needs are different. For instance, the improvement of Met+Cys provides value to the broiler chicken application, yet none for that of swine. Both applications would derive

considerable value from increased Lys and Thr levels (Bajjalieh, 2004). Soybean meal is the prevailing source of protein supplements in the poultry industry due to the fact that it is consistent in nutrition, is available year-round, and has high crude protein content. Broilers and turkeys consume roughly 44% of all the soybeans used in livestock feed in the USA, with layers using about 7% (Waldroup and Smith, 2001). It exceeds all other common plant protein feedstuffs in crude protein and digestible amino acid content (Waldroup and Smith, 2001). When blended with corn or grain sorghum meal, soybean meal provides growing poultry with a balanced diet of amino acids except for methionine. For swine nutrition, there are 10 essential amino acids: arginine (Arg), histidine (His), isoleucine (Iso), leucine (Leu), Lys, Met, phenylalanine (Phe), Thr, tryptophan (Trp), and valine (Val). The amino acids cysteine and tyrosine are considered semi-essential in swine diets, since both may be synthesized in sufficient amounts if the amino acids Met and Phe are present, respectively (Allee, 2005).

Legume seed storage proteins are categorized as either albumins or globulins based on their solubility patterns. Albumins are soluble in water, while globulins are extracted using dilute saline solutions. It has been found that most protein in soybean are globulins, which can be divided into the 7S vicilin-type and the 11S leguminin-type (Clarke and Wiseman, 2000). Glycinin and β-conglycinin represent the 11S and 7S fractions, respectively, based on their sedimentation properties (Danielsson, 1949). In combination, the glycinin and β-conglycinin fractions account for roughly 70% of the storage proteins in a soybean seed (Yaklich et al., 1999). Both have been found to be deficient in the sulfur-containing amino acids Cys and Met, with the 11S globulins higher in Met and Cys than the 7S type in general (Rajcan et al., 2005; Shewry et al., 1995). Methionine and Cys comprise only 3.0 to 4.5% of the 11S glycinin amino acid residues and less than 1% of the 7S β-conglycinin fractions (Nielsen et al., 1989). The

glycinins have been characterized extensively due to their nutritional value. Five major glycinin genes have been cloned (*Gy1, Gy2, Gy3, Gy4, and* Gy5) and grouped into two families based on the homology of their amino acid sequence. Group 1 includes *Gy1*, *Gy2*, and *Gy3*, and Group 2 includes *Gy4* and *Gy5* genes. The β-conglycinin fragment is a glycoprotein composed of three subunits: α, α', and β-subunit s. It is encoded by a multigene family containing 15 to 20 genes (Harada et al., 1989). The β-subunit lacks both Met and Cys, thus this subunit is primarily responsible for the low concentration of sulfur-containing amino acids in the β-conglycinin fraction (Sexton et al., 2002). It is likely that a soybean cultivar with a high 11S:7S ratio will have a higher concentration of the S-containing amino acids.

Methionine and Cys are the only two sulfur-containing amino acids and their production is tied directly to sulfur metabolism in the plant. Sulfur (S) is taken up from the soil as sulfate which is then distributed and mobilized in the plant. During the vegetative growth phase, sulfate reduction and incorporation of reduced S into amino acids takes place in developing leaves while these processes occur in developing seeds and pods during reproductive growth (Sexton et al., 2002). Two possible explanations for the deficiency of S-containing amino acids in soybean seed are is that these amino acids or other S compounds may not be supplied to the developing seed by the plant in a sufficient manner or by insufficient reduction of sulfate in the pods and seeds (Naeve and Shibles, 2005). It has been determined that accrual of the β-subunit is enhanced by excess nitrogen or by sulfur deficiency, thus low levels of S-containing amino acids in the seed may be due to the lack of reduced S compounds available to the seed (Naeve and Shibles, 2005; Paek et al., 1997). Paek et al (1997) speculated that the limitations in S-amino acid content may be due to three reasons: (i) the inability of the soybean plant to up-take sulfate rapidly,(ii) the inability to assimilate sulfate effectively, (iii) or the inability to mobilize S-amino

acids from vegetative tissue after mid-seed filling. In other words, high levels of Met, both *in vitro* and *in planta*, have been shown to inhibit the synthesis of the β-subunit of β-conglycinin, thereby increasing the relative quantity of Met and Cys in soybean seed storage proteins (Holowach et al., 1984).

There is great complexity in the biochemical pathways involved in the sequestering of Met and Cys in soybean seeds (Panthee et al., 2006b). The biosynthesis of Met has been reviewed previously (Hughes et al., 1999; Ravanel et al., 1998). There are two reactions which contribute to the regulation of Met biosynthesis. The first pertains to the shortened half-life of mRNA that encodes cystathionine γ-synthase, the enzyme that catalyzes the condensation of cysteine and *O*-phosphohomoserine to produce cystathionine due to elevated levels of Met (Chiba et al., 1999). In other words, the Met concentration may regulate the biosynthesis of the enzyme which is the catalyst in first step of Met biosynthesis. The other reaction involves Cys, which is a reactant in Met biosynthesis and is formed from *O*-acetylserine (OAS) and sulfide. The intracellular concentration of OAS may be responsible for the rate of Cys biosynthesis. Since Cys is a reactant in Met biosynthesis, the internal level of Cys may play an integral role in Met biosynthesis (Kim et al., 1999).

Plant breeders have been successful in increasing protein concentrations in soybean (Burton and Wilson, 1998; Weber and Fehr, 1970), though the concentrations of sulfur-containing amino acids in soybean cultivars have remained constant (Wilcox and Shibles, 2001). The difficulty in breeding for increased amino acid concentrations seems to stem from a lack of genetic variability for these traits (Krober, 1956). With that being said, breeders have assessed the amino acid quality of high-protein cultivars in the past in order to gain more of an understanding of this relationship. The average level of protein in soybean is approximately 400

g kg^{-1}. Improving protein has been successful, as cultivars with protein concentrations of 459 and 484 g kg^{-1} were developed using recurrent selection and backcross breeding, respectively (Brim and Burton, 1979; Wilcox and Cavins, 1995). 'Protana' (Weber and Fehr, 1970) and 'Prolina' (Burton and Wilson, 1998) are examples of soybean cultivars developed which have high protein. These cultivars were not readily adopted by growers due to their low yield potential. A strong negative correlation has been described between protein content and yield (Brim and Burton, 1979; Burton et al., 1982; Wilcox and Shibles, 2001). Some high-protein lines have been shown to improve nutritional value (Edwards 3rd et al., 2000), while others were unable to detect consistency in increased amino acid concentrations versus controls. Serretti (1994) found a high protein line with greater Cys concentration and one with lower Met concentration than the check genotype. Zarkadas (1993) found that increased protein was associated with reduced Met content while Yaklich (2001) found that both glycinin and β-conglycinin fractions were increased in high protein lines, with some lines having a greater proportion of glycinin polypeptides. This finding suggests that it is feasible to improve both the quantity and quality of soybean protein. It has also been shown that increasing the protein concentration results in just an increase in the β-conglycinin fraction of storage protein, thereby reducing overall protein quality (Nakasathien et al., 2000; Paek et al., 1997). Increased seed protein concentrations have been correlated with lower protein quality, particularly in the amino acid balances of Lys, Thr, and the sulfur-containing amino acids Met and Cys (Paek et al., 1997; Panthee et al., 2006b; Wilcox and Shibles, 2001; Wilson, 2004).

Currently, the relationship between protein quantity and quality is unclear, as previous results have shown. Overall, it is apparent that the development of cultivars with high protein quantity and quality has been elusive. Other breeding methodologies have been utilized to

address this problem as well. Using ethyl methanesulfonate (EMS) to mutate soybean seeds, Imsande (2001) was able to select several lines which overproduced Met and Cys at approximately a 20% higher level than the parental lines. There are also non-storage proteins within the soybean seed, such as protease inhibitors and seed lectin which may contribute up to 5% of the total seed protein. Protease inhibitors in soybean, which include the Bowman-Birk inhibitor (BBI) and the Kunitz trypsin inhibitor, reduce protein digestibility and are thus considered antinutritional compounds for animals and humans (Wilson, 1987). To deactivate these antinutritional components, raw soybean seed or soybean meal is heat-treated. This may actually reduce the nutritional quality of the soybean meal, since these protease inhibitors contain relatively high levels of essential sulfur-containing amino acids (Wilson, 1987).

Transgenic approaches adopted to increase sulfur-containing amino acids have entailed the introduction of transgenes from Brazil nut (*Bertholettia excelsa*) (Altenbach et al., 1987) and sunflower (*Helianthus annus*) (Kortt et al., 1991) encoding for proteins with extremely high levels of Met. Transformations were made in both tobacco (*Nicotiana tabacum* L.) (Altenbach et al., 1989) and soybean (Townsend and Thomas, 1994). Researchers have also expressed hydrophobic corn (*Zea mays* L.) proteins (δ-zeins) in soybean as a means of increasing Met and Cys (Kim, 2004). Both of these methods have resulted in little or no commercial improvement in amino acid expression in new soybean cultivars, as the Brazil nut protein was found to be allergenic and the δ-zein-transformed soybean line did not produce seed flour with significantly greater concentrations of sulfur-containing amino acids (Krishnan, 2008). Efforts have also been made to modify proteins already present within the soybean seed. Nielson (1989) identified a hypervariable region (HVR) between the Type 1 and Type 2 glycinin amino acid sequences. After inserting multiple Met residues in the HVR region of the *Gy4* gene and expressing the

modified gene in tobacco, the researchers failed to detected accumulated Met-enriched protein (Nielsen et al., 1995).

The development of molecular marker technology has also made it possible to create detailed linkage maps of soybean (Hyten et al., 2010; Song et al., 2004) and many other crop species (reference). This has facilitated the task of identifying chromosomal regions, quantitative trait loci (QTL), associated with particular traits. Molecular markers have previously been used to map chromosomal regions associated with protein (Brummer et al., 1997; Csanadi et al., 2001; Diers et al., 1992; Lee et al., 1996; Orf et al., 1999; Panthee, 2005; Sebolt et al., 2000) and amino acid concentration (Panthee et al., 2006a; Panthee et al., 2006b). Diers et al. (1992) used RFLP markers to map protein concentration in a F_2 population in which the high protein parent was *G. soja* accession. QTLs were identified on LG-I, -E, -F, and -G which explained between 12% and 42% of the variation and were associated with a 24 g kg^{-1} and 17 g kg^{-1}increase in protein concentration, respectively. Sebolt (2000) backcrossed these same alleles associated with increased protein on LG-I and LG-E into an elite background. Only the protein QTL on LG-I was detected in this study. Brummer et al. (1997) evaluated eight different soybean populations using RFLP markers and identified QTLs on LG-A2, -B2, -C1, -D1, -E, -F, -G, -H, and -I conditioning protein concentration.

There have been several studies aimed at elucidating the genetic factors underlying amino acid concentration in soybean. Using 101 F_6-derived recombinant inbred lines (RILs), Panthee et al. (2006b) identified QTL associated with Cys (chr 1, 13, and 18), Met (chr 13, 18, and 7), and Met+Cys (chr 13 and 7) concentration. Panthee et al. (2006a) also identified genomic regions associated with Lys (chr 1, 15, and 18) and Thr (chr 5, 2, 9, and 19). In a similar study, Panthee et al.(2004) used the same RILs to map QTL associated with the 7S and 11S fractions of

soybean storage proteins. Since the glycinin fraction contains higher levels of S-containing amino acids than the β-conglycinin fraction, the identification of genomic regions governing these storage protein constituents would be beneficial in marker assisted selection (MAS) schemes for improved S-containing amino acids. They found three QTL for glycinin (chr 17, 20, and 19) and two for β-conglycinin concentration (chr 17 and 16).

At the present time, not a single commercial cultivar of soybean with the FAO standard total sulfur containing amino acids has been developed. With continued improvements in breeding technologies in conjunction with continued elucidation of QTLs, it is foreseeable in the near future that a soybean cultivar with high yield, protein, and improved levels of essential sulfur-containing amino acids will be developed.

Genotype × Environment Interaction

With the world population set to reach 10 billion by 2050 it will be necessary to increase agricultural production to meet such needs. Yield stability has been of major concern to plant breeders as well as national and international programs with aims of maximizing the yield potential of certain regions, while minimizing crop failures or extremely low yields in poor growing years (Annicchiarico, 2002). The key to keeping pace with population growth will be improving the efficiency of agricultural production practices and resource usage (Kang, 1997). The efforts of plant breeders will be essential in advancing agricultural production and farming systems in poor growing regions (Sleper et al., 1991). Understanding and exploiting G×E interactions has become an integral part of these efforts. Genotype × environment interaction refers to differential genotypic responses in different environments; these interactions affect all living organisms, not just plants. The differential expression of genotypes across environments minimizes the association between the genotypic and phenotypic values, resulting in differences

between genotypes which are not the same across all environments. The G×E interaction can influence decisions in a breeding program such as the breath of adaptation of a new cultivar, selection of a location for early generation testing, the type of environment (ie. stress vs non-stress) used in testing, and the geographic distribution for testing prior to final release (Kang, 1997). The G×E interaction and stability of quality traits in soybean are of great interest to breeders and growers especially in markets where premium pricing is utilized. Lin and Binns (1994) classified G×E interactions into three groups in terms of areas of research interest. The goal of some groups is to find a model, such as principal component analysis (Zobel et al., 1988), to explain the structure of G×E interaction or to attempt to predict it. The second type of research interest is of the quantitative genetic type, where the goal is to estimate the size of the interaction as a variance for use in the prediction of genetic improvement in selection. Lastly, the plant breeders utilize G×E interaction in selecting superior stable cultivars and determining and recommending growing environments.

In the context of G×E, the term genotype refers to a cultivar such as a pure line, clone, or open-pollinated populations. In this context, the term does not necessarily pertain to the collection of genes which make up an individual, as it is usually defined. Phenotype refers to the outward appearance or the traits of an individual at a physical, morphological, anatomical, or biological level, and is the result of both genetic and non-genetic factors. Breeders seek to improve quantitative traits of crop plants by selecting genotypes based on their phenotypic performance. Genotypic expression across a range of environments confers a range of phenotypes and selection only takes advantage of those factors which are genetic in nature (Comstock and Moll, 1963; Kang, 1997). Environment refers to the total of circumstances surrounding a genotype, or the set of climatic, soil, biotic (pests and diseases) and management

conditions in an individual experiment or trial carried out at a location during one year (annuals) or several years (perennials). Since the genetic makeup of an individual does not change from environment to environment, any phenotypic variation for a specific genotype is due to the environment. Allard and Bradshaw (1964) defined the variation in environments as either predictable or unpredictable. The predictable type refers to permanent conditions that occur systematically such as general climate, soil type, and day length or those which are fixed under human control such as planting date, sowing density, and harvest system. The overall crop performance may be the best indicator of this type of variation and the presence of genotype × location interactions is indicative of different environments while genotype × treatment interactions indicate that the treatments provide differing growing conditions (Allard, 1964; Fehr, 1987). Unpredictable factors refer to those which undergo random fluctuations such as rainfall, temperature, and relative humidity. This involves genotype × year and genotype × location × year interactions phenomena which cannot be predicted in advance, thereby making it very different from breeding for the aforementioned predictable type of environmental variation.

Genotype × environment interactions may be grouped into two broad categories: crossover and noncrossover interactions. Crossover interaction of genotypes occurs when cultivars change ranks across environments. In this case, the genotype favored by selection will differ between environments. These are the most important type of G×E effects targeted by breeders as they decrease the heritability of a trait and hinder genotypic evaluation (Burton, 1987). Non-crossover interactions represent changes in the magnitude of the difference between two genotypes and may mean that cultivars are genetically heterogeneous but test environments are homogeneous, or genotypes are genetically homogeneous but environments are heterogeneous. With non-crossover interactions, estimates of heritability and predicted

phenotypic response will decrease but less than for crossover interactions while the genotype favored by selection will not change (Kang, 1997). In this situation, all identical genotypes grown in constant (ideal) environments would rank similarly.

Stress is the main causal factor in G×E interactions. Stress represents a physiological response to the effect of a negative environmental factor and is present when any factor is present at any other level other than the optimum. These may include nutrients, toxic elements, salts in the soil solution, atmospheric gases, light of differing wavelengths, mechanical stimuli, gravity, wounding, pests, pathogens, and symbionts (Crispeels, 1994). Biotic factors which cause G×E interaction may include plant pathogens and pests, nutrient uptake ability, competition between genotypes, tolerance to herbicides, allelopathy, and water-, nutrient- and radation-use efficiency (Kang, 1997). Abiotic stresses are also a culprit in G×E interactions, and may include atmospheric pollutants, soil stresses, temperature, water, and tillage operations (Blum, 1988; Clark and Duncan, 1993; Specht and Laing, 1993). Environments with contrasting levels of one major stress have frequently shown high levels of G×E interaction (Ceccarelli, 1989).

The two primary methods by which breeders assess the prominence and quantitative nature of G×E are by investigating the components of variance from the analysis of variance (ANOVA) or by conducting stability evaluations. To detect interactions between genotypes and the environment, the genotypes are grown over a range of environments composed of multiple locations and/or years. If G×E interaction is absent, then all genotypes should perform similarly across environments and therefore the total variation is explained by just the main effects of environments and genotypes (Chahal and Gosal, 2002). The ANOVA allows for the calculation of variance components for each source of variation in the model, thus allowing breeders to gain

an understanding of the most prominent sources of variation. Numerous studies have been undertaken which use variance components to explore G×E interactions (Erickson et al., 1982; Johnson et al., 1955; Kwon and Torrie, 1964; Zhe et al., 2010). The general consensus among breeders is that G × E interaction is associated more with quantitative traits than with those of a qualitative nature (Hoisington et al., 1982; Langrdige and Griffing, 1959). Quantitative traits exhibit continuous variation due to polygenic gene action and/or differences among environments, thus such traits tend to have low heritability.

Genotype × environment interaction has been explored in soybean for a number of quantitative traits including yield, important agronomic traits, and compositional traits. Johnson et al. (1962) evaluated yield, height, seed weight, and oil in F_3 lines in the F_4 and F_5 generation and found differing levels of G×E interaction based on variance component analysis. The researchers indicated the need for testing over multiple environments due to the fact that genetic variability was reduced by 71% when genotypes were grown in only one location in one year as opposed to multiple years and locations. Two soybean populations were evaluated for G×E in the F_3, F_4, and F_5 generations by Kwon and Torrie (1964). The line × year variance component estimates were larger than either the line × location or the line × location × year variance components for yield, seed weight, lodging, days to flowering, and percent oil. The genotype variance component for protein was greater than the interaction components, but less than the error variance. Similarly, Erikson (1982) found that the genotype × location × year variance component for protein was larger than either two-way component, but less than the genotype variance component.

The concept of stability has many aspects and may be viewed in a number of ways. Allard and Bradshaw (1964) emphasized that the stability of a genotype refers to facets of the

phenotype like yield and quality which are economically important, not just general constancy of all phenotypes across environments. In other words, stability is concerned only with traits of interest; other traits may vary among environments and are of no consequence. Cultivars which can change their genotypic or phenotypic response as environmental conditions vary and give high and stable returns on an economic level have been termed "well-buffered" or homeostatic (Allard, 1964; Lewontin, 1957). Allard and Bradshaw (1964) described two general ways in which a genotype achieves stability. First, a cultivar may be made up of a number of different genotypes which are adapted to different environments, termed "population buffering". "individual buffering" refers to the adaptedness of individuals themselves to a range of environments. To this end, it has been theorized that the genetic structure of plant material may have an effect on the extent to which G × E interaction is present (Schutz, 1971; Walker, 1978). Pure-lines and single-cross hybrids, which are highly homogeneous and in the case of the pure-line and clonal cultivar, homozygous, have been shown to interact more with the environment than open-pollinated cultivars or mixtures of pure-lines due to the fact that they have fewer adaptative genes due to their genetic structure and are therefore more susceptible to environmental variation (Becker and Leon, 1988).

Two types of stability, static and dynamic, have been previously described (Becker and Leon, 1988). Static stability refers to the situation in which the performance of a genotype for some trait remains unchanged when grown in multiple environments (ie. its variance among environments is equal to zero). This type of stability is useful for traits like quality traits, resistance to pathogens and diseases, and those conferring resistance to stress where it is imperative that levels be maintained (Becker and Leon, 1988). Dynamic stability, on the other hand occurs when a genotype's performance corresponds to the predicted response of each

environment. Quantitative traits like yield are usually evaluated for stability in terms of dynamic stability, as breeders prefer to identify environments which produce high-yielding lines. Genotypes which deviate significantly from the general reaction of the genotype across locations are viewed as being unstable in this situation. Becker (1981) coined static stability as agronomic stability, while referring to dynamic stability as the biological concept of stability.

Stability parameters may be grouped into four different categories (Kang, 1997). The following describes a few of the more popular stability parameters within each grouping, but is not all-inclusive. Type 1 stability refers to a cultivar that has a small variance over a range of environments. This type of stability can be assessed by simply determining the variance of a cultivar across environments or the coefficient of variation of genotypes across environments. The Finlay and Wilkinson (1963) regression coefficient (b = 0 is considered stable) is also a form of Type 1 stability. Type 2 stability considers genotypes with performance which is parallel to the mean of all genotypes in the test to be stable. This type of stability includes Plaisted's (1960) variance component analysis for G×E interaction, the Eberhart and Russell (1966) stability parameter (b = 1 is considered stable), Wricke's (1962) ecovalence, and Shukla's (1972) stability variance (σ_i^2). The residual mean square of deviation from the regression variance (δ_i^2) is the second part of the Eberhart and Russell (1966) stability parameter. This defines a cultivar to be stable when this value is small, also called Type 3 stability. Lastly, a genotype is considered stable if the year (or seeding date) mean squares within locations is small (Lin and Binns, 1988). This type of stability can only be detected when the experiment includes genotype × location × time (year or seeding date) interaction.

An integral component in establishment of markets for soybean cultivars with value-added traits is the determination of how traits respond to changes in environment. The most

pragmatic and cost-effective situation would be one in which cultivars express wide adaptation over a range of growing environments. Since multi-location and multi-year testing is a costly and time-consuming endeavor, it is necessary that breeders determine the optimum combination of replicates and environments to sufficiently measure the genetic value of a genotype. Increased replications at each location function to improve the precision of measurements and the power to detect differences between genotypes within the location. Conversely, the presence of G×E interaction between the test environment and the breeder's base population of environments negates the effectiveness of more replicates at a single location. Thus, more precision would be gained in discriminating between genotypes by adding additional testing environments as opposed to replications (Bernardo, 2002). The addition of environments is associated with increased resource input including labor, land, and supplies (Kang, 1997). The best location or combination of locations should provide a measure of the relative potential of genotypes over the target population of environments and maximize genetic variation, and in turn, response to selection (Allen and Rasmusson, 1978). With that being said, there appears to be a trade-off between precision and resource allocation. Schutz and Bernard (1967) estimated the interaction variance for the Soybean Uniform Tests in Maturity Groups 0 to IV, VI, and VII and found that testing yield in more than 20 environments did not reduce Fisher's Least Significant Difference (LSD) significantly and that 10 environments would be suitable for testing.

Due to the importance of G×E interactions, crop genotypes are usually assessed in multi-environment trials prior to their release as cultivars. Environmental effects within the base population of environments themselves are generally not of concern to breeders. The genotypic main effects and genotype × environment interactions, on the other hand, provide relevant

information. Genotypic main effects are of no consequence to the breeder when G×E interaction is present, except in cases where a significant interaction is considered in determining the significance of the genotypic effects (random model). Plant breeders must deal with G×E interaction in some practical manner (Kang, 1997). Three ways of dealing with G×E interaction have been expressed in the literature (Bernardo, 2002; Eisemann et al., 1990) include ignoring these interactions, avoiding them, or exploiting them in breeding objectives. The amount of G×E interaction greatly impacts how breeding programs allocate resources as multiple breeding programs may be needed in an area where there is prevalent G interaction.

If genotypic ranks change drastically between environments, a breeder either has to develop separate populations for each location, or select genotypes which perform well over all environments. In the first scenario, the breeder will see greater genetic gains, but increased costs. The second yields less genetic gain but is also less expensive (McKeand et al., 1990). In order to effectively and efficiently improve a quantitative trait the breeder must quantify the amount and nature of the G×E interaction.

References

Allard R.W.a.A.D.B. (1964) Implications of genotype-environmental interactions in applied plant breeding. Crop Science 4:503-508.

Allee G. (2005) Protein and Amino Acid Sources for Swine Diets. National Swine Nutrition Guide:30-36.

Allen F., Rasmusson R. (1978) Optimal Environments for Yield Testing1. Crop Science 18:747.

Altenbach S.B., Pearson K.W., Leung F.W., Sun S.S.M. (1987) Cloning and sequence analysis of a cDNA encoding a Brazil nut protein exceptionally rich in methionine. Plant molecular biology 8:239-250.

Altenbach S.B., Pearson K.W., Meeker G., Staraci L.C., Sun S.S.M. (1989) Enhancement of the methionine content of seed proteins by the expression of a chimeric gene encoding a methionine-rich protein in transgenic plants. Plant molecular biology 13:513-522.

Annicchiarico P. (2002) Genotype x environment interactions: challenges and opportunities for plant breeding and cultivar recommendations Food & Agriculture Org.

Bajjalieh N. (2004) Proteins from oilseeds, FAO.

Becker H. (1981) Correlations among some statistical measures of phenotypic stability. Euphytica 30:835-840.

Becker H., Leon J. (1988) Stability analysis in plant breeding. Plant Breeding 101:1-23.

Bernardo R. (2002) Breeding for quantitative traits in plants Stemma Press. Woodbury, MN.

Blum A. (1988) Plant breeding for stress environments.

Brim C., Burton J. (1979) Recurrent selection in soybeans. II. Selection for increased percent protein in seeds. Crop Science 19:494.

Brummer E., Graef G., Orf J., Wilcox J., Shoemaker R. (1997) Mapping QTL for seed protein and oil content in eight soybean populations. Crop Science 37:370-377.

Burton J. (1987) Quantitative genetics: results relevant to soybean breeding. Agronomy (USA).

Burton J., Wilson T. (1998) Registration of 'Prolina'soybean. Crop Science 39:294.

Burton J., Purcell A., Walter Jr W. (1982) Methionine concentration in soybean protein from populations selected for increased percent protein. Crop Sci 22:430-432.

Ceccarelli S. (1989) Wide adaptation: how wide? Euphytica 40:197-205.

Chahal G., Gosal S. (2002) Principles and procedures of plant breeding: biotechnological and conventional approaches Alpha Science Intl Ltd.

Chiba Y., Ishikawa M., Kijima F., Tyson R.H., Kim J., Yamamoto A., Nambara E., Leustek T., Wallsgrove R.M., Naito S. (1999) Evidence for autoregulation of cystathionine -synthase mRNA stability in Arabidopsis. Science 286:1371.

Clark R., Duncan R. (1993) Selection of plants to tolerate soil salinity, acidity and mineral deficiencies. Int. Crop Sci 1:371-379.

Clarke E., Wiseman J. (2000) Developments in plant breeding for improved nutritional quality of soya beans I. Protein and amino acid content. The Journal of Agricultural Science 134:111-124.

Comstock R., Moll R. (1963) Genotype-environment interactions, National Academies. pp. 164.

Crispeels M. (1994) Introduction to Signal Transduction in Plants: a Collection of Updates. Am. Soc. Plant Physiologists.

Csanadi G., Vollmann J., Stift G., Lelley T. (2001) Seed quality QTLs identified in a molecular map of early maturing soybean. TAG Theoretical and Applied Genetics 103:912-919.

Danielsson C. (1949) Seed globulins of the Gramineae and Leguminosae. Biochemical Journal 44:387.

Diers B., Keim P., Fehr W., Shoemaker R. (1992) RFLP analysis of soybean seed protein and oil content. TAG Theoretical and Applied Genetics 83:608-612.

Durham D. (2003) The United Soybean Board's better bean initiative: Building United States soybean competitiveness from the inside out.

Eberhart S., Russell W. (1966) Stability parameters for comparing varieties. Crop Sci 6:36-40.

Edwards 3rd H., Douglas M., Parsons C., Baker D. (2000) Protein and energy evaluation of soybean meals processed from genetically modified high-protein soybeans. Poultry Science 79:525.

Eisemann R., Cooper M., Woodruff D. (1990) Beyond the analytical methodology: better interpretation and exploitation of Genotype-by-Enviromnent interaction in breeding. Genotype-by-environment interaction and plant breeding:108–117.

Erickson L.R., Beversdorf W.D., Ball S.T. (1982) Genotype × Environment Interactions for Protein in Glycine max × Glycine soja crosses1. Crop Sci. 22:1099-1101. DOI: 10.2135/cropsci1982.0011183X002200060003x.

Fehr W.R. (1987) Principles of cultivar development. Volume 1. Theory and technique Macmillan publishing company.

Finlay K., Wilkinson G. (1963) The analysis of adaptation in a plant-breeding programme. Australian Journal of Agricultural Research 14:742-754.

Friedman M., Brandon D. (2001) Nutritional and Health Benefits of Soy Proteins†. J. Agric. Food Chem 49:1069-1086.

George A.A., De Lumen B.O. (1991) A novel methionine-rich protein in soybean seed: Identification, amino acid composition, and N-terminal sequence. Journal of Agricultural and Food Chemistry 39:224-227.

Harada J.J., Barker S.J., Goldberg R.B. (1989) Soybean [beta]-Conglycinin Genes Are Clustered in Several DNA Regions and Are Regulated by Transcriptional and Posttranscriptional Processes. The Plant Cell Online 1:415.

Hoisington D.A., Neuffer M.G., Walbot V. (1982) Disease lesion mimics in maize I. Effect of genetic background, temperature, developmental age and wounding on necrotic spot formation with *Les 1*. . Developmental Biology 93:381-388.

Holowach L.P., Thompson J.F., Madison J.T. (1984) Effects of exogenous methionine on storage protein composition of soybean cotyledons cultured in vitro. Plant physiology 74:576.

Hughes C.A., Gebhardt J.S., Reuss A., Matthews B.F. (1999) Identification and expression of a cDNA encoding cystathionine [gamma]-synthase in soybean. Plant Science 146:69-79.

Hyten D.L., Choi I.Y., Song Q., Specht J.E., Carter Jr T.E., Shoemaker R.C., Hwang E.Y., Matukumalli L.K., Cregan P.B. (2010) A High Density Integrated Genetic Linkage Map of Soybean and the Development of a 1536 Universal Soy Linkage Panel for Quantitative Trait Locus Mapping. Crop Science 50:960-968.

Imsande J. (2001) Selection of soybean mutants with increased concentrations of seed methionine and cysteine. Crop Science 41:510-515.

Johnson H., Bernard R. (1962) Soybean genetics and breeding. Advances in Agronomy 14:149-221.

Johnson H., Robinson H., Comstock R. (1955) Genotypic and phenotypic correlations in soybeans and their implications in selection. Agron. j 47:477-483.

Kang M. (1997) Using genotype-by-environment interaction for crop cultivar development. Advances in Agronomy 62:199-252.

Kerley M., Allee G. (2003) Modifications in soybean seed composition to enhance animal feed use and value: Moving from a dietary ingredient to a functional dietary component.

Kim H., Hirai M.Y., Hayashi H., Chino M., Naito S., Fujiwara T. (1999) Role of O-acetyl-L-serine in the coordinated regulation of the expression of a soybean seed storage-protein gene by sulfur and nitrogen nutrition. Planta 209:282-289.

Kim W.S., and H.B. Krishnan. (2004) Expression of an 11 kDa methionine-rich delta-zein in transgenic soybean results in the formation of two types of novel protein bodies in transitional cells situated between the vascular tissue and storage parenchyma cells. . J.Plant Biotech. 2:199-210.

Kortt A., Caldwell J., Lilley G., Higgins T. (1991) Amino acid and cDNA sequences of a methionine-rich 2S protein from sunflower seed (Helianthus annuus L.). European journal of biochemistry/FEBS 195:329.

Krishnan H. (2008) Improving the Sulfur-Containing Amino Acids of Soybean to Enhance its Nutritional Value in Animal Feed. Sulfur: A Missing Link Between Soils, Crops, and Nutrition:235.

Krober O. (1956) Nutritive Quality of Crops, Methionine Content of Soybeans as Influenced by Location and Season. Journal of Agricultural and Food Chemistry 4:254-257.

Kwon S., Torrie J. (1964) Heritability and interrelationship among traits of two soybean populations. Crop Sci 4:196-198.

Langrdige J., Griffing B. (1959) A study of high temperature lesions in *Arabidopsis thaliana*. Molecular and General Genetics 194:15-23.

Lee S., Bailey M., Mian M., Carter T., Shipe E., Ashley D., Parrott W., Hussey R., Boerma H. (1996) RFLP loci associated with soybean seed protein and oil content across populations and locations. TAG Theoretical and Applied Genetics 93:649-657.

Lewontin R.C. (1957) The adaptations of populations to varying environments, Cold Spring Harbor Laboratory Press. pp. 395.

Lin C., Binns M. (1988) A superiority measure of cultivar performance for cultivar× location data. Canadian Journal of Plant Science 68:193-198.

Lin C., Binns M. (1994) Concepts and methods for analyzing regional trial data for cultivar and location selection.

McKeand S., Li B., Hatcher A., Weir R. (1990) Stability parameter estimates for stem volume for loblolly pine families growing in different regions in the southeastern United States. Forest science 36:10-17.

Naeve S.L.S., Shibles R.M. (2005) Distribution and mobilization of sulfur during soybean reproduction. Crop Science 45:2540.

Nakasathien S., Israel D., Wilson R., Kwanyuen P. (2000) Regulation of seed protein concentration in soybean by supra-optimal nitrogen supply.

Nielsen N., Jung R., Nam Y.W., Beaman T., Oliveira L., Bassüner R. (1995) Synthesis and assembly of 11S globulins. Journal of plant physiology 145:641-647.

Nielsen N.C., Dickinson C.D., Cho T.J., Thanh V.H., Scallon B.J., Fischer R.L., Sims T.L., Drews G.N., Goldberg R.B. (1989) Characterization of the glycinin gene family in soybean. The Plant Cell Online 1:313.

Orf J., Chase K., Jarvik T., Mansur L., Cregan P., Adler F., Lark K. (1999) Genetics of soybean agronomic traits: I. Comparison of three related recombinant inbred populations. Crop Science 39:1642.

Paek N., Imsande J., Shoemaker R., Shibles R. (1997) Nutritional control of soybean seed storage protein. Crop Science 37:498-503.

Panthee D. (2005) Genetic mapping of quantitative trait loci conditioning protein concentration and quality, and other seed characteristics in soybean [Glycine max (L.) Merrill].

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

Panthee D., Kwanyuen P., Sams C., West D., Saxton A., Pantalone V. (2004) Quantitative trait loci for -conglycinin (7S) and glycinin (11S) fractions of soybean storage protein. Journal of the American Oil Chemists' Society 81:1005-1012.

Panthee D., Pantalone V., Sams C., Saxton A., West D., Orf J., Killam A. (2006b) Quantitative trait loci controlling sulfur containing amino acids, methionine and cysteine, in soybean seeds. TAG Theoretical and Applied Genetics 112:546-553.

Plaisted R. (1960) A shorter method for evaluating the ability of selections to yield consistently over locations. American Journal of Potato Research 37:166-172.

Rajcan I., Hou G., Weir A. (2005) Advances in breeding of seed-quality traits in soybean. Journal of Crop Improvement 14:145-174.

Ravanel S., Gakière B., Job D., Douce R. (1998) The specific features of methionine biosynthesis and metabolism in plants. Proceedings of the National Academy of Sciences of the United States of America 95:7805.

Schutz W. (1967) Genotype Environment Interactions in the Regional Testing of Soybean Strains1. Crop Science 7:125.

Schutz W. (1971) Inter-Genotypic Competition in Soybeans. III. An Evaluation of Stability in Multiline Mixtures1. Crop Science 11:684.

Sebolt A., Shoemaker R., Diers B. (2000) Analysis of a quantitative trait locus allele from wild soybean that increases seed protein concentration in soybean. Crop Science 40:1438.

Serretti C., Schapaugh Jr W., Leffel R. (1994) Amino acid profile of high seed protein soybean. Crop Science 34:207.

Sexton P., Paek N., Naeve S., Shibles R. (2002) Sulfur metabolism and protein quality of soybean. Journal of Crop Production 5:285-308.

Shewry P., Napier J., Tatham A. (1995) Seed storage proteins: structures and biosynthesis. The Plant Cell Online 7:945.

Shukla G. (1972) Some statistical aspects of partitioning genotype-environmental components of variability. Heredity 29:237.

Sleper D.A., Barker T.C., Bramel-Cox P. (1991) Plant breeding and sustainable agriculture: considerations for objectives and methods proceedings of a symposium. CSSA special publication (USA).

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

Specht J., Laing D. (1993) Selection for tolerance to abiotic stresses–discussion. International Crop Science I. Crop Science Society of America, Madison, Wisconsin:381–382.

Townsend J., Thomas L. (1994) Factors which influence the Agrobacterium-mediated transformation of soybean. J. Cell Biochem. Suppl. 18A 78.

Waldroup P.W., Smith D.R.K. (2001) Soybean Use-Poultry.

Walker A. (1978) Yield Stability of Soybean Mixtures and Multiple Pure Stands1. Crop Science 18:719.

Weber C., Fehr W. (1970) Registration of Provar soybeans. Crop Sci 10:728.

Wilcox J., Cavins J. (1995) Backcrossing high seed protein to a soybean cultivar. Crop Science 35:1036.

Wilcox J., Shibles R. (2001) Interrelationships among seed quality attributes in soybean. Crop Science 41:11.

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

Wilson R. (2004) Seed composition. Soybeans: improvement, production, and uses 3:621–677.

Wricke G. (1962) Uber eine Methode zur Erfassung der okologischen Streubreite in Feldversuchen. Z. Pflanzencuhtg 47:92-96.

Yaklich R. (2001) [beta]-Conglycinin and Glycinin in High-Protein Soybean Seeds. J. Agric. Food Chem 49:729-735.

Yaklich R., Helm R., Cockrell G., Herman E. (1999) Analysis of the distribution of the major soybean seed allergens in a core collection of Glycine max accessions. Crop Science 39:1444.

Zarkadas C.G., Yu Z., Voldeng H.D., Minero-Amador A. (1993) Assessment of the protein quality of a new high-protein soybean cultivar by amino acid analysis. Journal of Agricultural and Food Chemistry 41:616-623.

Zhe Y., Lauer J.G., Borges R., de Leon N. (2010) Effects of Genotype × Environment Interaction on Agronomic Traits in Soybean. Crop Sci. 50:696-702. DOI: 10.2135/cropsci2008.12.0742.

Zobel R.W., Wright M.J., Gauch Jr H.G. (1988) Statistical analysis of a yield trial. Agronomy journal (USA).

CHAPTER 3

MAPPING QTL FOR SEED PROTEIN AND AMINO ACIDS IN THE BENNING × DANBAEKKONG SOYBEAN POPULATION[1]

[1] Warrington, C.V., D.H. Hyten, P.B. Cregan, J.H. Orf, A.S. Killam, N. Bajjalieh, H. Abdel-Haleem, and H.R. Boerma.. To be submitted to *Crop Science*

Abstract

Soybean [*Glycine max* (L.) Merr.] is the primary source of quality protein in feed formulations for the domestic swine, poultry, beef, and dairy industries. As a sole dietary source of protein, soybean is deficient in the amino acids lysine (Lys), threonine (Thr), methionine (Met), and cysteine (Cys) for poultry and swine. Increasing these amino acids would benefit the feed industry. The objective of this study was to identify quantitative trait loci (QTL) associated with crude protein (cp), Lys/cp, Thr/cp, Met/cp, and Cys/cp and Met+Cys/cp in a population of 140 F_5-derived RILs from a 'Benning' × 'Danbaekkong' cross. The 140 RILs and check cultivars were grown in five southern USA environments. A seed sample from each RIL was analyzed by near-infrared reflectance spectroscopy to determine amino acid concentration as a fraction of cp and amino acid concentrations. Each RIL was genotyped with 421 polymorphic markers (98 simple sequence repeat markers and 323 single nucleotide polymorphism markers). Putative QTL were detected using single factor ANOVA and composite interval mapping (CIM). A large-effect QTL on chr 20 inherited from Danbaekkong which explained 55% of the phenotypic variance was detected for crude protein based on CIM. This QTL was also detected for Lys/cp, Thr/cp, Met/cp, Cys/cp and Met+Cys/cp, but the Danbaekkong allele resulted in reduced levels of these amino acids. Based on CIM, three other QTL were detected for crude protein on chr 14, 15, and 17, two for Lys/cp on chr 8 and 20, and three for Thr/cp were detected on chr 9, 17, and 20. Four QTL were found on chr 6, 9, 10, and 20 for Met/cp, and one QTL was detected for Cys/cp on Chr 10. Transgressive segregation in this population was identified for crude protein, Met/cp, and Met+Cys/cp. This study provides important information concerning the relationship between crude protein and levels of essential amino acids and may allow for the improvement of these traits in soybean using marker-assisted selection (MAS).

Introduction

As the world population increases, so too will the demand for animal protein. It is projected that global meat production and consumption will increase from 233 million tonnes in 2000 to 300 million t by the year 2020 (Annicchiarico, 2002). Soybean [*Glycine max* (L.) Merr.] is the world's leading oilseed crop and provider of high-quality protein meal. Soybean meal, a by-product of the oil extraction, drives the soybean market mainly due to its use as a source of amino acids in livestock and poultry rations. The animal feed industry uses roughly 77% of soybean meal as a source of protein and amino acids (Kerley and Allee, 2003). With this in mind its importance to the U.S. agriculture industry cannot be overstated.

Legume seed proteins are categorized as either albumins or globulins based on their solubility patterns. It has been found that most protein in soybean are globulins, which can be divided into the 7S vicilin-type and the 11S leguminin-type (Clarke and Wiseman, 2000). Glycinin and β-conglycinin represent the 11S and 7S fractions, respectively, based on their sedimentation properties (Danielsson, 1949). In combination, the glycinin and β-conglycinin fractions account for roughly 70% of the storage proteins in a soybean seed (Yaklich et al., 1999). Both have been found to be deficient in the sulfur-containing amino acids cysteine (Cys) and methionine (Met), with the 11S globulins generally higher than the 7S type (Rajcan et al., 2005; Shewry et al., 1995). It has been found that Met and Cys comprise 3.0 to 4.5% of the 11S glycinin amino acid residues and less than 1.0% of the 7S β-conglycinin fractions (Nielsen et al., 1989). The β-conglycinin fragment is composed of three subunits: α-, α'-, and β-subunit. The β-subunit lacks both Met and Cys, thus this subunit is primarily responsible for the low concentration of sulfur-containing amino acids in the β-conglycinin fraction. It is likely that a

soybean cultivar with a high 11S:7S ratio will have a higher concentration of the S-containing amino acids.

Protein, *per* se, is not of great importance in terms of animal nutrition. Yet the balance and composition of the amino acid constituents which comprise the protein is likely the most crucial nutritional aspect of meal rations. The major function of protein in nutrition is to supply adequate amounts of required amino acids (Friedman and Brandon, 2001). As a sole dietary source of protein, soybean is deficient in the amino acids Met, Cys, threonine (Thr), and lysine (Lys). Each of these amino acids is considered essential, as monogastric animals (e.g., swine and poultry) cannot synthesize these amino acids and therefore each must be obtained solely from the diet. Any deficiency in the amino acid balance must be supplemented in the diet at additional costs to the animal producer. To overcome deficiencies, poultry and swine growers supplement soybean-based rations with synthetically-produced amino acids, a process which costs approximately $100 million annually (Imsande, 2001). Clarke and Wiseman (2000) speculated that a 10% increase in Lys, Met, and Thr concentrations would yield a $4.5 to 9.5, $2.7, and $5.9/T increase in commercial meal value, respectively. Moreover, according to George and de Lumen (1991), Met supplementation may cause additional problems such as leaching during soybean meal processing and bacterial degradation leading to the formation of undesirable volatile sulfides. Therefore, the development of soybean cultivars with enhanced amino acid balance would increase their economic value along the entire soybean value chain, from grower to end-user, and reduce any negative environmental effects associated with supplementation.

Due to the aforementioned issues, it is not surprising that the development of soybean cultivars with increased concentrations of essential amino acids has been an objective in the soybean breeding community for some time. The negative correlation between protein and yield

has undermined attempts to release cultivars with higher levels of essential amino acids (Wilson, 2004). Until recently, the cost of wet lab techniques necessary to evaluate amino acid concentrations made it costly for breeders to thoroughly evaluate protein and amino acid concentrations in large seed samples.

There is great complexity in the biochemical pathways involved in the sequestering of Met and Cys in soybean seeds (Panthee et al., 2006b). Increases in seed protein concentration have also been correlated with lower protein quality, particularly in the amino acid balances of Lys, Thr, and the sulfur-containing amino acids Met and Cys (Paek et al., 1997; Panthee et al., 2006b; Wilcox and Shibles, 2001; Wilson, 2004). The United Soybean Board's Better Bean Initiative (BBI) included as one of its major research goals in 2002 that soybean breeders develop soybean cultivars in the USA with increased seed protein and improved seed protein quality so as to better compete with foreign producers that can devote more agricultural land to soybean production (Sallstrom, 2002). One of the primary traits targeted by the BBI is increased Met+Cys concentration, while increased levels of Lys and Thr represent secondary goals of the initiative. The value of increased levels of these amino acids is application driven, as swine and poultry needs are different. For instance, the improvement of Met+Cys provides value to the broiler chicken application, yet none for that of swine. Both applications would derive considerable value from increased Lys and Thr levels (Bajjalieh, 2004).

Efforts made in the past to address amino acid content have mostly been aimed at improving the sulfur-containing amino acids. Conventional plant breeding, the introduction of transgenes, raising the expression of endogenous Met-rich proteins, and amending soil nutrients are methods by which researchers have sought to address this objective.

Plant breeders have been successful in increasing protein concentrations in soybean (Burton and Wilson, 1998; Weber and Fehr, 1970) , though the concentrations of sulfur-containing amino acids have remained the same (Wilcox and Shibles, 2001). Difficulty in breeding for increased amino acid concentrations stem from the lack of genetic variability for these traits (Krober, 1956). Some high-protein lines have been shown to improve nutritional value (Edwards 3rd et al., 2000), but others were unable to detect consistency in increased amino acid concentrations versus controls (Serretti et al., 1994). Yaklich (2001) found that high protein soybean lines had increased glycinin and β-conglycinin fractions, with some lines having a greater proportion of glycinin polypeptides, suggesting that it is feasible to improve both the quantity and quality of soybean protein. Using ethyl methanesulfonate (EMS) to mutate soybean seeds, Imsande (2001) was able to select several lines which overproduced Met and Cys at approximately a 20% higher level than the parental lines.

Transgenic approaches to increase sulfur-containing amino acids have entailed the introduction of transgenes from Brazil nut (*Bertholettia excelsa*) (Altenbach et al., 1987) and sunflower (*Helianthus annus*) (Kortt et al., 1991)encoding for proteins with extremely high levels of Met. Transformations were made in both tobacco (*Nicotiana tabacum* L.) (Altenbach et al., 1989) and soybean (Townsend and Thomas, 1994) . Researchers have also expressed hydrophobic corn (*Zea mays* L.) proteins (δ-zeins) in soybean as a means of increasing Met and Cys (Kim, 2004). Both of these methods have resulted in little or no improvement in amino acid expression in newly released soybean cultivars, as the Brazil nut protein was found to be allergenic and the δ-zein-transformed soybeans did not produce seed flour with significantly greater concentrations of sulfur-containing amino acids (Krishnan, 2008). Falco (1995) was able to stably produce soybean lines with increased lysine through transformation, but lines with

greater than 15% lysine produced seed with wrinkled seed coats and poor germination. Efforts have also been made to modify proteins already present within the soybean seed. Nielson (Nielsen, 1990) identified a hypervariable region (HVR) between the Type 1 and Type 2 glycinin amino acid sequences. After inserting multiple Met residues in the HVR region of the *Gy4* gene and expressing the modified gene in tobacco, the researchers failed to detected accumulated Met-enriched protein (Nielsen et al., 1995).

The nutrient amendment approach to improve sulfur-containing amino acids included regulating nitrogen and sulfur in the soil (Imsande and Schmidt, 1998; Sexton et al., 1998). In these studies, the researchers assessed seed quality of soybean with respect to two major seed storage proteins, the β-conglycinin (7S) and glycinin (11S) fractions, following differential levels of nitrogen and sulfur soil supplementation. Their respective goals were to increase the glycinin fraction of seed storage proteins due to the fact that 11S contains higher Met concentrations than 7S. In comparison to the breeding and biotechnological methods aimed at improving this amino acid quality, this approach appears to be unsustainable (Panthee et al., 2006b).

There have been few studies aimed at elucidating the genetic factors underlying amino acid concentration in soybean. Using 101 F_6-derived recombinant inbred lines (RILs), Panthee et al. (Panthee et al., 2006b) identified QTL associated with Cys (chr 1, 13, and 18), Met (chr 13, 18, and 7), and Met+Cys (chr 13 and 7) concentration. Panthee et al. (2006a) also identified genomic regions associated with Lys (chr 1, 15, and 18) and Thr (chr 2, 5, 9, and 19). In a similar study, Panthee et al. (2004) used the same RILs to map QTL associated with the 7S and 11S fractions of soybean storage proteins. Since the glycinin fraction contains higher levels of S-containing amino acids than the β-conglycinin fraction, the identification of genomic regions governing these storage protein constituents would be beneficial in marker assisted selection

(MAS) regimes for improved S-containing amino acids. They found three QTL for glycinin (chr 17, 19, and 20) and two for β-conglycinin concentration (chr 16 and 17).

In order to efficiently develop soybean cultivars with improved amino acid profiles, the genetic basis of amino acid content should be explored thereby allowing for the selection of individual components conditioning improved protein quality. The objectives of this study were to identify QTL associated with crude protein, Lys/cp, Thr/cp, Met/cp, Cys/cp, and Met+Cys/cp in a RIL population created from a cross of 'Benning' and 'Danbaekkong'.

Materials and Methods

Plant Material

A population of 140 F_5-derived recombinant-inbred lines (RILs) was developed from a cross of Benning (PI595645) (Boerma,1997) × Danbaekkong (PI619083) (Kim, 1996). The parents were chosen based on their disparate protein levels, with Benning averaging approximately 42% and Danbaekkong at 51% on a dry-weight basis. Benning is a high-yielding maturity group VII cultivar adapted to the southeastern USA and Danbaekkong is a South Korean maturity group IV tofu cultivar.

From the original cross, seeds from individual F_1 plants were grown in the greenhouse and seed from individual plants were bulked. The F_2 plants were grown at the Univ. of Georgia Plant Sciences Farm near Watkinsville, GA. Seeds from individual F_2 plants were advanced to the F_5 generation in Athens, GA and Puerto Rico using a modified single seed descent (Brim, 1966). The F_3 and F_4 generations were grown in Puerto Rico and the F_5 generation was grown at the Univ. of Georgia Plant Sciences Farm. At maturity individual F_5 plants were single-plant threshed to create F_5-derived recombinant inbred lines (RILs). Approximately 200 RILs were grown in 2003 and 150 RILs were selected for uniform maturity.

In 2005 and 2006, 150 RILs were planted at the Univ. of Georgia Plant Sciences Farm near Athens, GA and also planted in Bay, AR, Stuttgart, AR, and Kinston, NC in 2006. The 150 RILs were sub-divided into three sets of 50 RILs based on their relative maturity. Danbaekkong and three check cultivars, 'NCRoy', 'AG6202', and 'Boggs-RR' were included in each set. For each set the experimental design was a randomized complete block with two replications. Each set was also randomized within a single replication.

The experiment was planted in Athens in an Appling loamy coarse sand soil type on 19 May 2005 and on a Cecil coarse sandy loam (fine, kaolinitic, thermic Typic Kanhapludults) soil type on 22 May 2006 and were irrigated. The experimental unit in Athens was a 2-row plot that was 7-m long with 76-cm between rows and was seeded with approximately 27 seeds $^{-1}$m row. At maturity, all plots were end-trimmed to a final row length of 3.66 m and the plots were harvested by plot combine. The experiments were planted in a mixture of Mhoon (Fine-silty, mixed, superactive, nonacid, thermic Fluvaquentic Endoaquepts) and Dundee (Fine-silty, mixed, active, thermic Typic Endoaqualf) fine sandy loam soil type in Bay, AR on 13 June 2006 and were irrigated. The experimental unit was a 2-row plot with 76-cm between rows. In Stuttgart, AR, the experiments were planted in a Stuttgart silt loam (Fine, smectitic, thermic Albaquultic Hapludalfsoil) soil type on 26 May 2006 and were irrigated. The experimental unit was a 2-row plot with 76-cm between rows. The Kinston, NC experiments were planted on 19 June 2006 in a Portsmouth soil type (loam with ~4% organic matter) (Fine-loamy over sandy or sandy-skeletal, mixed, semiactive, thermic Typic Umbraquults) and were not irrigated. The experimental unit was a 1-row plot harvested from within 3 planted rows spaced 96-cm apart. The plots were end-trimmed to 4.26 m prior harvesting the middle row of each plot.

Phenotypic Data

Soybean samples were scanned by near infrared (NIR) spectrometry at the Univ. of Minnesota's Soybean Breeding Laboratory. The samples were analyzed in the laboratory of Dr. Jim Orf (University of Minnesota, St. Paul, MN), in collaboration with Dr. Nick Bajjalieh (Integrative Nutrition Inc., Decatur, IL), and were reported in g kg^{-1} on a moisture-free basis. The crude protein and amino acid analyses were conducted on 25-g whole seed samples with near-infrared reflectance (NIR). Whole soybean samples were first ground using a Perten LM 3600 grinder and then scanned on a FOSS 6500 NIR Instrument. NIR spectra from the FOSS 6500 were predicted using ISIPredict Software version 1.10.2.4842. Each amino acid sample was corrected as a percentage of overall crude protein content (also in g kg^{-1})

Genotyping

Each RIL was genotyped with 421 polymorphic markers, including 98 simple sequence repeats (SSR) and 323 single nucleotide polymorphisms (SNP). For the SSR marker analysis, DNA from 140 greenhouse-grown RILs (10 leave samples) was extracted from unexpanded trifoliolate leaves using a modified CTAB (Hexadecyltrimethylammonium acid) procedure previously described by Keim et al. (2006). For PCR amplification (32 cycles, 94ºC for 1 min, 94ºC for 30 sec, 46ºC for 30 sec, 68ºC for 30 sec, and held at 10ºC after final cycle), reaction mixtures contained 20 ng of genomic DNA, 0.5 µM of forward and reverse primers (Grant et al., 2002), 2 mM of each dNTP, 2.5 mM Mg^{2+}, 1X PCR buffer (Promega Corp., Madison, WI), and 0.5 units *Taq* polymerase (Promega Corp., Madison, WI) in a total volume of 10 µl. The separation of PCR amplicons was conducted using 4.8% polyacrylamide gels run on either an ABI PRISM 377 DNA Sequencer (PE ABI, Foster City, CA) or using capillary gel electrophoresis using an ABI 3730 DNA analyzer (Applied Biosystems, Foster City, CA).

For PCR product fragment length analysis on the ABI PRISM 377, samples were prepared for electrophoresis by combining 3 µl of PCR product, 2 µl formamide, 0.75 µl loading buffer, and 0.30 µl GENESCAN-500 ROX DNA size standard (PE Applied Biosystems, Foster City, CA). For ABI 3730 genotyping, 8 µl of a master mix of ROX size standard, water, and formamide was added to 2 µl of DNA. Each sample was denatured for 5 min at 95ºC, and then loaded into the gel or capillary system. Gels were scored visually based on marker size data from each parent to determine the SSR marker genotype of each line.

The RILs were also genotyped with single nucleotide polymorphism (SNP) markers using the Illumina GoldenGate Assay (Hyten et al., 2008a). DNA was extracted from a 10 leaf sample and processed to contain 50 µl of DNA at a 200 ng/µl concentration. The samples were then sent to the USDA Beltsville Agricultural Research Center (USDA-ARS) in Beltsville, MD, where a total of 1,536 SNP markers were assayed on each RIL genotype using the Universal Soybean Linkage Panel 1.0 (USLP 1.0) (Hyten et al., 2010b), using the GoldenGate® assay and analyzed on the Illumina BeadStation 500G (Illumina, San Diego, CA) (Hyten et al., 2008b). Of the 150 RILs genotyped, only 140 were used in the QTL analysis due to aberrant segregation ratios most likely due to a seed mixture.

Data Analyses

For linkage map construction, a total of 421 markers, 323 SNP and 98 SSR markers were analyzed using MapDisto v 1.7 (Lorieux, 2007) mapping software. A more stringent LOD threshold of 3.0 was used to identify initial linkage groups, followed by a more conservative LOD score of 1.5 to group each LG individually. The recombination fraction setting used was classical, based on Martin (2006) selected. The Kosambi (1944) mapping function was used in

order to address interference. Based on recombination frequencies, 28 linkage groups (LG) were created, which are a representation of the 20 haploid chromosomes in the soybean genome.

Single-factor analysis of variance (SF-ANOVA) was used to detect associations between markers and traits using (QTL) Cartographer V2.5_006 ($P \leq 0.001$) (Wang et al., 2007). Each marker is considered a factor with two levels (homozygous Benning or homozygous Danbaekkong) and the phenotype (protein or specific amino acid) as the dependent variable. Composite interval mapping method (CIM) was employed to detect QTLs and estimate the magnitude of their effects (Jansen and Stam, 1994) using Model 6 of the Zmapqtl program module. A series of 1000 permutations was run to determine the experiment-wise significant level at P = 0.05 of LOD for each trait (Churchill and Doerge, 1994). The genome was scanned at 2-cM intervals and the window size was set at 10 cM. Cofactors were chosen using the forward-backward method of stepwise regression. Putative QTL were further analyzed using multiple regression until only significant markers were retained in the model ($P \leq 0.01$) using the STEPWISE selection criteria (SAS, 2003). All possible two-way interactions between significant markers (P = 0.01) were evaluated for significance (P = 0.01) by ANOVA using PROC GLM (SAS, 2003) to evaluate the presence of epistasis for each trait.

Variance-component heritability estimates were calculated on an entry-mean basis (Nyquist and Baker, 1991) using the following equation:

$$H^2 = \frac{\sigma_g^2}{\sigma_g^2 + (\sigma_g^2/e) + (\sigma^2/re)}$$

where H^2 represents broad-sense heritability, σ^2 is genotypic variance, σ^2_{ge} is genotype × environment variance, σ^2 is error variance, r is the number of replications, and e is the number of environments. Restricted maximum likelihood (REML) was used to generate components of

variance and covariance for calculating heritabilities and genetic correlations. Genetic correlations were based on the formula ((Falconer and Mackay, 1996):

$$r_G = \frac{Cov_{xy}}{\sqrt{\sigma_x^2 \sigma_y^2}}$$

where r_G is the genetic correlation, Cov_{xy} is the covariance of trait x and trait y, x symbolizes trait x, y symbolizes trait y, and σ^2 is the genetic variance.

Results

The male parent, Danbaekkong, and three elite cultivars, NC Roy (Burton et al., 2005) , AG6202, and Boggs-RR (Boerma et al., 2000) were used as a checks in all environments. The maternal parent, the maturity group VII 'Benning', was not suitable as a check due to its late maturity (Relative Maturity of 7.8) when compared to the RILs evaluated in this study. Danbaekkong is a late maturity group IV cultivar and the RILs averaged 47 days in maturity after 31 August compared to 50, 44, 48, and 50 for AG6202, Boggs-RR, and NC Roy respectively (data not shown).

Across the five environments, Danbaekkong averaged 510 g kg^{-1} seed protein content while the three elite checks averaged 433 g kg^{-1} (Table 2.1). The 140 RILs ranged from 425 to 507 g kg^{-1} seed protein content and averaged 468 g kg^{-1}. Based on the comparison of the mean protein content of Danbaekong and the RIL with the highest protein level, there was no transgressive segregation for this trait. The variance component heritability for protein on a five-environment mean basis was 0.93.

In this manuscript the amino acid data are presented as the amount of a specific amino acid per kg of crude protein (cp). The mean Lys value for Danbaekkong was 61.2 g kg^{-1} cp compared to 64.8 g kg^{-1} cp for the mean of the three elite checks (Table 2.1). The mean Lys/cp content for the RILs was 63.3 g kg^{-1} and they ranged from 61.2 to 65.0 g kg^{-1}. The distribution

of the Lys/cp contents of the140 RILs seemed to approximate a normal distribution (Fig.2.1). None of the RILs contained lower Lys content than Danbaekkong or a higher value than the mean of the three elite check cultivars. The variance component heritability for Lys/cp based on the selection of Lys means across five environments (0.69) was somewhat lower than the heritability for crude protein content.

The Danbaekkong parent had a value Thr value of 34.4 g kg^{-1} cp while the elite checks averaged 37.9 g kg^{-1}. The RILs averaged 36.1 g kg^{-1} for this trait and ranged from 33.8 g kg^{-1} to 38.0g kg^{-1} (Table 2.1). There was no significant (P = 0.05) transgressive segregation (Fig. 2; Table 1). Variance component heritability for this trait was 0.86 based on the afformentioned selection criteria.

For Met/cp, both Danbaekkong and the checks were very similar, as Danbaekkong was 14.2 g kg^{-1} and the elite checks were 14.4 g kg^{-1} (Table 2.1). The mean of the RILs was 14.2 g kg^{-1}, equal to the Danbaekkong parent. Transgressive segregation was evident, as the lines varied from 13.8 to 14.7 g kg^{-1} (Fig 2.3). Heritability for Met/cp, 0.45, was much lower than crude protein or Thr/cp.

The three checks averaged 15.3 g kg^{-1} cp Cys compared to 15.7 g kg^{-1} for Danbaekkong (Table 2.1). The mean of RILs was equal to the check means, and the highest value for any RIL was 0.5 g kg^{-1} greater than the Danbaekkong parent (Fig. 2.4). The 140 RILs ranged from 14.7 to 16.2 g kg^{-1}. The variance component heritability for Cys/cp (0.59) was slightly higher than that of Met/cp.

Danbaekkong had a value of 29.9 g kg^{-1} for Met+Cys/cp and the elite checks averaged 29.7 g kg^{-1} (Table 2.1). The RILs averaged 29.6 g kg^{-1} and revealed transgressive segregation, as

values for this trait ranged from 28.5 to 30.8 g kg^{-1} (Fig. 2.5). Heritability for this trait (0.53) was intermediate between both of its constituents, Met/cp and Cys/cp, alone.

Based on the RIL means across the five environments crude protein content was negatively correlated with Lys/cp (r – 0.63**), Thr/cp (r = -0.85**), Met/cp (r = -0.19*), and Cys/cp (r = -0.16). Genotypic correlations, which indicate the direction and magnitude of correlated responses to selection (Falconer and Mackay, 1996) were also calculated. Genotypic correlations for crude protein versus Lys/cp and Thr/cp were -0.83 and -0.91, respectively. The genotypic correlation coefficients for crude protein versus Met/cp were -0.36 and -0.06 for crude protein versus Cys/cp.

Phenotypic correlations based on the five environment means for the 140 RILs between the various amino acids were positive, ranging from 0.41 (Thr/cp vs. Cys/cp) to 0.78 (Lys/cp vs. Thr/cp) (Table 2.2). Met/cp was found to be highly correlated with both Lys/cp (r = 0.70**) and Cys/cp (r = 0.57**). Genetic correlations ranged from 0.17 (Lys/cp vs. Cys/cp) to 0.92 (Lys/cp vs Thr/cp).

Figure 2.6 shows markers on linkage groups created from the Benning × Danbaekkong population aligned with markers from the consensus map 4.0 (Hyten et al., 2010a). The 421 polymorphic markers (98 SSRs and 323 SNPs) mapped in the RIL population provided broad coverage of most of the 20 soybean linkage groups with only a few exceptions. An exception was chr 7, which contained only six linked markers (Fig 2.6). On the average the Benning × Danbaekkong linkage map contained a marker approximately every 5 to 6 cM, although gaps greater than 40 cM (based on the consensus map positions) occurred in chr2 (Lg-D1b), chr4 (Lg-C1), chr6 (Lg-C2), chr14 (Lg-B2). Chromosome 18 (Lg-G) had the most markers (5 SSR and 70 SNP markers). Some chromosomes had sparse marker coverage in certain regions which caused

the mapping software to split single LGs into two subgroups (e.g. Chr2, 4, 7, 11, 14, 16, 19, and 20). The marker order on the LGs and the subgroups is in general agreement with that of the integrated genetic linkage map (Consensus Map 4.0) (Hyten et al., 2010a). The 28 chromosomes covered roughly 1124 cM of the genome. Although a fairly large number of markers were used in this study, many SNP markers were clustered in regions of the genome, therefore reducing genome coverage.

Given the amino acids were expressed as a percent of crude protein, it was critical to identify the protein QTLs in this population. Based on SF-ANOVA at $P \leq 0.01$, protein QTL on seven chromosomes were found (Table 2.3). These QTL explained from 4 to 53% of the variation in crude protein content. The QTL located on chr 20 (Lg-I) accounted for over 8 times the variation of the next largest protein QTL (chr 6, chr 7). The allele for increased protein at the QTL on chr 20 was inherited from Danbaekkong and when homozygous resulted in over 28 g kg^{-1} greater protein than the Benning allele. Other protein QTLs were found on chr 10, chr 13, chr 14, and chr 15. The alleles for increased protein were inherited from Danbaekkong at all QTL with the exception of the QTL on chr 13. Composite interval mapping (CIM) identified QTL on chr 14, chr 15, chr 17, and chr 20 (Table 2.3; Fig. 2.6f,g,h,i). The analysis indicated these QTL accounted for 5 to 55% of the variation in protein content. As found by SF-ANOVA, the QTL on chr 20 near BARC-061899 accounted for more variation in crude protein content than the other three QTL combined. The significant markers identified in Tables 2.3 and 2.4 were analyzed using the STEPWISE selection criteria of PROC REG (SAS, 2003). For crude protein, BARC-061899 (R^2 = 60%), BARC-042781 (R^2 = 4%), and BARC-018353 (R^2 = 3%) remained in the multiple regression model and explained 67% of the variation for crude protein combined. Based on the heritability value of crude protein (0.93), the markers were able to explain 72% of

the genotypic variation for this trait. No significant epistasis was detected between significant markers for crude protein.

The SF-ANOVA identified three chromosomes associated with Lys/cp concentration ($P \leq 0.01$) on chr 5, chr 15, and chr 20 (Table 2.4). The locus on chr 20 accounted for the majority of the variation for this trait ($R^2 = 47\%$) while the other two were minor in their effects. At each QTL for this trait the allele for increased Lys/cp was inherited from Benning. The effect of Benning alleles in the homozygous state at the BARC-061899 QTL resulted in an increase in Lys/cp of 1.2 g kg^{-1} while the other QTL increased Lys/cp from 0.31 to 0.49 g kg^{-1} in the same allelic state. The CIM analysis for Lys/cp identified two intervals, one on chr 8 and a second on chr 20 (Fig 2.6 c, i). The alleles on chr 20 explained roughly the same amount of variation for this trait as was detected using SF-ANOVA ($R^2 = 48\%$). The locus on chr 8 accounted for only 6% of the variation and was also identified at a less stringent significance level in the SF-ANOVA ($P \leq 0.05$). BARC-016899 ($R^2 = 49\%$), Satt231 ($R^2 = 4\%$), and BARC-055265 ($R^2 = 2\%$) were retained in the STEPWISE multiple regression analysis, explaining 55% of variation for Lys/cp. These markers explained 80% of the genotypic variation for Lys/cp based on the heritability estimate for this trait (0.69). No significant epistasis was detected between significant markers for Lys/cp.

Five QTL were discovered for Thr/cp using SF-ANOVA at $P \leq 0.01$. Similar to crude protein and Lys/cp, the QTL with the greatest impact was on chr 20 at the BARC-061899 locus, and accounted for 51% of the variation (Table 2.5). The remaining four QTL accounted for 6% or less of the variation for this trait (chr 4, chr 7, chr 9, and chr 10). The Benning parent was the donor of positive alleles for Thr/cp at all loci and when homozygous at the BARC-061899 locus increased it by 1.3 g kg^{-1}. The next greatest effect at a homozygous Benning locus was 0.4 g kg^{-1}

for Satt478 on chr 10. The CIM identified four QTL (Chr. 1, 9, 17, and 20), two of which were not identified in the SF-ANOVA at the $P \leq 0.01$ significance level (BARC-035219 on chr. 1($P = 0.05$) and Satt256 on chr. 17) (Fig. 2.6 a,d,h,i). BARC-061899 on chr 20 accounted for roughly 9 times more phenotypic variation in Thr/cp than the other QTL discovered with CIM (Chr 1, 9, and 17). BARC-016899 ($R^2 = 54\%$), BARC-035219 ($R^2 = 4\%$), BARC-048619 ($R^2 = 3\%$), and Satt256 ($R^2 = 3\%$) were retained in the STEPWISE multiple regression analysis, explaining 63% of variation for Thr/cp. Based on the heritability estimate of 0.86 for Thr/cp the markers retained in the model explained 73% of the genotypic variation for this trait. No significant epistasis was detected between significant markers for this trait.

For Met/cp SF-ANOVA detected a QTL on chr. 20 ($R^2 = 12\%$), but it was 9 cM distal to the BARC-061899 for Thr/cp and Lys/cp at the BARC-020713 locus (Table 2.4). Two other QTL were detected on chr 9 and chr10 at BARC-042449 ($R^2 = 8\%$) and Satt592 ($R^2 = 11\%$), respectively. Positive alleles were inherited from Benning in each case, accounting for increases in Met/cp between 0.11 and 0.16 g kg^{-1}. QTL on chr 6, chr 9, chr 10, and chr 20 were significant for CIM, with BARC-020713 explaining the most variation of the four ($R^2 = 20\%$) (Table 2.4; Fig 2.6 b,d,e,i). The remaining three loci explained 8, 9, and 14% of the variation for the QTL on chr 6, 9, and 10, respectively. The QTL on chr 6 was detected in the SF-ANOVA at a less stringent probability ($P \leq 0.05$). Four markers, BARC-020713 ($R^2 = 14\%$), Satt592 ($R^2 = 9\%$), BARC-042449 ($R^2 = 5\%$), and BARC-055889 ($R^2 = 4\%$) explained approximately 32% of the variation for Met/cp based on the multiple regression analysis. These four markers explain 71% of the genotypic variation based on the heritability estimate of 0.45 for this trait. These markers did not interact significantly.

For Cys/cp, SF-ANOVA detected QTL at four loci (Table 2.4). Satt592 on chr 10 explained the most variation (R^2 = 13%), while the remaining three each explained 6% or less of the variation. Two QTL inherited their positive alleles from Benning, Satt592 and BARC-020713, resulting in increased Cys concentration by 0.20 and 0.14 g kg^{-1}, respectively, when in the homozygous state, respectively. The positive alleles were inherited from Danbaekkong at the chr 6 and chr 14 QTL and each provided an 0.11 g kg^{-1} increase in Cys/cp. The only QTL identified by CIM was Satt592 on chr 10, which explained 10% of the variation for Cys/cp (Fig 2.6 e). The multiple regression analysis for Cys/cp identified three markers explaining the variation for this trait. BARC-020713 (R^2 = 49%), Satt592 (R^2 = 4%), and BARC-048543 (R^2 = 2%) were retained in the model and explained 55% of the phenotypic variation for this trait and 93% of the genotypic variation based on the heritability estimate for Cys/cp of 0.59. No epistaticinteraction was detected between significant markers for Cys/cp.

Similar to Met/cp and Cys/cp, the two loci with the largest effects for Met+Cys/cp were Satt592 on chr 10 and BARC-020713 on chr 20, which explained 15% and 10% of the phenotypic variation, respectively. At these two loci, the positive alleles were inherited from Danbaekkong. Three QTL with positive alleles inherited from Benning were detected on chr 6, chr 14, and chr 18 and explained between 4 and 5% of the variation in Met+Cys/cp. Increases in Met+Cys/cp of 0.34 and 0.30 would be expected when the Benning alleles are homozygous at the QTL identified on chr 10 and chr 20. The three minor QTL for Met+Cys/cp which inherited their positive alleles from Danbaekkong would result in 0.16 to 0.22 g kg^{-1} improvement in Met+Cys/cp concentration. The two loci on chr 10 and chr20 were significant in the interval mapping analysis, explaining roughly the same amount of phenotypic variation (R^2 = 11-12%) (Fig 2.6 e,i). The BARC-020713 locus (R^2 = 8%) was the only marker retained in the

STEPWISE selection procedure in the multiple regression analysis for Met+Cys/cp. Based on the heritability for this trait (0.53%), this QTL only explained 15% of the genotypic variation for this trait. This may be due the fact that the effects of each QTL for each individual amino acid, Met or Cys, may be confounded in this analysis when the values were combined. No significant epistasis was detected between markers for Met+Cys/cp.

Discussion

Soybean recombinant inbred lines were recovered with crude protein greater than 500 g kg^{-1} which is near the maximum detected historically for soybean accessions in the germplasm collection (Yaklich, 2001). This finding was not surprising as the Danbaekkong parent averages 510 g $kg^{-1}$of protein. As has been the major impediment in the efforts to develop high yielding, high protein cultivars, we detected a strong negative correlation (r = -0.50) between protein and seed yield (Warrington, 2011). The amino acid values as a percent of total crude protein detected in this study are similar to those set forth by the National Research Council (1994; 1998) for poultry and swine nutrition, though far from the trait end points proposed by the United Soybean Board's Better Bean Initiative (Bajjalieh, 2004; Sallstrom, 2002).

The large effect QTL conditioning crude protein found on chr 20 (Lg-I) at BARC-061899 has been identified previously in other populations. QTL conditioning protein were also detected by Brummer et al. (1997), Diers et al. (1992), Sebolt et al. (2000),Chung et al. (2003), and Mansur et al. (1993) on chr 20. Nichols (2006) fine-mapped the seed protein QTL on chr 20 using two sets of backcross lines to the region between SSR marker Satt239 and AFLP marker ACG9b. In our study, we detected a SNP marker (BARC_016899) within this interval which is highly significant for crude protein, Lys/cp, Thr/cp, Met/cp, and Met+Cys/cp. This interval

corresponds to the major protein QTL cqPRO-003 which was previously identified and confirmed (Nichols et al., 2006).

Few studies have undertaken the task of elucidating the genomic regions associated with amino acid concentration in soybean. Panthee et al.(2006a; 2006b) used SSR markers to map amino acid concentration (dry weight basis) in a soybean population of 101 F_6-derived recombinant inbred lines (RILs). They identified QTL associated with Cys (chr 1, 13, and 18), Met (chr 13, 18, and 7), and Met+Cys (chr 13 and 7) concentration (Panthee et al., 2006b). Panthee et al. (2006b) also identified genomic regions associated with Lys (chr 1, 15, and 18) and Thr (chr 2, 5, 9, and 19). For the sulfur-containing amino acids, no QTL detected in these studies were found on the same chromosomes as our mapping population. A QTL for Lys was detected roughly 20 cM upstream from a QTL we detected on chr 15. In addition, Panthee (2006a) reported a QTL on chr 9 for Thr. We report one on this chromosome as well, but around 40 cM downstream. Panthee et al. (2004) used the same RILs to map QTL associated with the 7S and 11S fractions of soybean storage proteins. Since the glycinin fraction contains higher levels of S-containing amino acids than the β-conglycinin fraction, the identification of genomic regions governing these storage protein constituents would be beneficial in MAS regimes for improved S-containing amino acids. They reported three QTL for glycinin (chr 17, 19, and 20) and two for β-conglycinin concentration (chr 16 and 17). The QTL detected for glycinin on chr 20 was 45 cM downstream from the QTL we detected for crude protein on the same chromosome. We found QTL on chr 17 for crude protein and Thr/cp. The crude protein QTL and the Thr QTL are approximately 8 cM and 43 cM from the β-conglycinin QTL detected in their study, respectively.

Soybase (Grant et al., 2002) (www.soybase.org, verified 15 March 2011) reports a number of protein and protein-related QTLwithin 10cM upstream or downstream of those detected in our study. As previously mentioned, a number of protein QTL have been previously detected in the same region of chr 20 as found in our study. Nine QTL associated with protein are reported in Soybase near our QTL. These include Prot 1-1, Prot 1-2, Prot 1-3, Prot 1-4, Prot 3-12, Prot 11-1, Prot 15-1, Prot 17-1, and the previously mentioned cqPro-003 QTL. Two QTL (Prot 4-5 and Prot 4-6) were detected on chr 15 approximately 6 cM from our QTL for crude protein. Interestingly, three QTL relevant to protein quality (Glycinin 1-1, Acidic fraction 1-1, and Conglycinin 1-1) are reported in Soybase roughly 8 cM from the QTL detected in this region for crude protein and Thr/cp in the current study. A protein quality QTL (Acidic fraction 1-3) and a protein QTL (Prot 13-4) was reported 2 and 4 cM from the QTL we detected for Thr/cp and Met/cp, respectively. A QTL for protein in Soybase (Prot 24-1) was mapped to an identical region of chr 6 in our study.

Since the parental genotypes should be fixed at nearly all allelic loci, the transgressive segregation present for some of the traits is likely due to the complementary action of additive alleles that are dispersed between the parental lines, a byproduct of recombination (Rieseberg et al., 1999). Transgressive segregation is proof that there are effects from QTL alleles inherited from both parental genotypes. Otherwise, if all the positive alleles came from just one of the parents, the highest value for the progeny would be equal to that of the highest parent. This is the case for Lys/cp and Thr/cp, where there was no significant transgressive segregation. In this study, the amount of transgressive segregants can only be interpreted in regard to the Danbaekkong parent since Benning was not grown in the experiment. Our findings show that Danbaekkong possesses nearly all of the positive alleles for crude protein (Table 2.3), while only

Benning alleles lead to increased levels of Lys/cp, Thr/cp, and Met/cp (Table 2.4). On the other hand, alleles from both parents can lead to improved phenotypes for Cys/cp and Met+Cys/cp (Table 2.4).

QTL conditioning crude protein, Lys/cp, Thr/cp, Met/cp, Cys/cp, and Met+Cys/cp were detected using single factor analysis of variance (SF-ANOVA) and composite interval mapping (CIM). It is evident from our results that the QTL on chr 20 (Lg-I) has a great impact on both protein quantity and quality. The variation explained (R^2 = 55%) by the crude protein QTL identified at the SNP marker BARC-061899 is the highest reported in the literature to date. The genotyping of 421 polymorphic SNP markers in this population provided increased precision compared to the SSR-based maps utilized previously to map these traits. When the allele at BARC-061899 is inherited from Danbaekkong, this QTL also reduces Lys/cp, Thr/cp, Met/cp, Cys/cp, and Met+Cys/cp. The fact that positive alleles for protein quality are not inherited along with the Danbaekkong allele for higher protein quantity is crucial to our understanding of how to best develop genotypes with improved amino acid profiles. The aim is to introgress alleles which improve protein quality without sacrificing protein quantity. In nearly all cases, the positive alleles for amino acid concentrations were inherited from the Benning parent. When the Danbaekkong allele on chr 20 at BARC-061899 is homozygous, crude protein is increased by approximately 28.2 g kg^{-1}in this population of RILs. RILs homozygous for the same allele will average 1.2 and 1.3 g kg^{-1}less in Lys/cp and Thr/cp, respectively. BARC-020713 on chr 20, which is significant for Met/cp, Cys/cp, and Met+Cys/cp is 9.1 cM upstream from BARC-061899. Based on both SF-ANOVA and CIM, this is the same QTL as BARC-061899. The result of inheriting the Danbaekkong allele at this QTL(s) is a reduction in Met/cp by 0.16 g kg^{-1}, Cys/cp by 0.14 g kg^{-1}, and Met+Cys/cp by 0.30 g kg^{-1}.

Yates et al. (2006) found that backcross-derived lines containing this same high-protein allele within a different genetic background produced seeds with lower levels of Thr and Lys, but no change in the levels of Met or Cys. This generally fits the expectation given the large effect on both Lys and Thr shown in our results. Based on the aforementioned allelic arrays and their affects on crude protein and the amino acid profile, it is not feasible to select only for the major crude protein QTL on chr 20 and improve protein quality. By selecting for the Danbaekkong allele on chr 20 and for either Danbaekkong or Benning alleles at QTL on other chromosomes which affect protein quality, breeders may be able to improve protein and maintain protein quality concurrently. The QTL detected for Lys/cp and Thr/cp on other chromosomes do not increase values for these two traits even half as much as the QTL at BARC-061899 on chr 20. On the other hand, concentrations of the sulfur-containing amino acids may be improved by introgressing Danbaekkong alleles at Satt592 on chr 10 and Benning alleles at QTL on chr 6, 14, and 18 while still increasing the level of protein with the Danbaekkong allele at the chr 20 QTL. The increase in Met+Cys provided by this locus is actually greater (2a = 0.34) than that on chr 20 (2a = 0.30) (Table 2.8). Another approach would be to maintain crude protein, while increasing sulfur-containing amino acids. This could be accomplished by selecting for the Benning allele at the chr 6, 14, 18, and 20 QTL and the Danbaekkong allele at the QTL for Met+Cys/cp on chr 10.

Developing a clear understanding the relationship between protein concentration and quality has been difficult. It is evident that the crude protein QTL detected in our study on chr 20 also plays some role in the sequestration of the other amino acids within the seed. It is clear from both phenotypic and genotypic correlations that an increase in crude protein results in decreased values of these amino acids (Table 2.2). The negative correlation between crude protein and Thr is especially strong. In terms of the sulfur-containing amino acids, it is known

that the synthesis of storage protein within the developing soybean seed is sensitive to Met concentration, in that the presence of Met during this process prevents the synthesis of low quality proteins; in other words if the plants are grown in sulfur deficient soils then poor quality seed storage proteins will be synthesized (Sexton et al., 2002).

In plants, Met, Lys, and Thr are part of the aspartate family of amino acids, thus synthesized from the same precursor, aspartate. Aspartate is the carboxylate ion, or ester, of the non-essential amino acid aspartic acid and is vital in the biosynthesis of these amino acids (Shen et al., 2002). Therefore, it is not surprising that these three amino acids were correlated. It is of interest to look at aspartate kinase and aspartate semialdehyde dehydrogenase, as they are first two enzymes which function in the pathway and could therefore be responsible for increased or reduced levels of each of these amino acids.

Wilcox and Shibles (2001) found that Met and Cys levels remained constant even when protein was increased. Cycles of recurrent selection which increased protein from 438 to 474 g kg^{-1} over six cycles did not significantly change Met concentration (Burton et al., 1982). On the other hand, it has also been shown that increasing the protein concentration results in an increase in the β-conglycinin fraction of storage protein, thereby reducing overall protein quality (Nakasathien et al., 2000; Paek et al., 1997). Serretti (1994) found a high protein line with greater Cys concentration and one with lower Met concentration than the check genotype. Findings by Paek et al. (1997) suggested that soybean seed incorporates as much sulfur amino acids as is available and then produces the poorer β-subunit based on the availability of nitrogen. This same group speculated that the limitations in S-amino acid content may be due to three reasons: (i) the inability of the soybean plant to up-take sulfate rapidly, (ii) to assimilate sulfate effectively, or (iii) the inability to mobilize S-amino acids from vegetative tissue after mid-seed

filling (Paek et al., 1997). Grabau (1986) was able to produce seed with 23 and 31% increases in Met and Cys concentrations, respectively, by providing soybean plants with a reduced form of sulfur (Met) during seed filling. From these findings, it is apparent that when sulfur is present seeds will accumulate 11S proteins and no β-subunits of the 7S fraction (Paek et al., 2000; Sexton et al., 1998).

This study reinforces the fact that breeding efforts for soybean quality need not focus completely on increasing protein concentration. Simply increasing crude protein may not increase essential amino acid concentrations. Mapping both crude protein and amino acids concurrently within the same population allows for a more precise understanding of the interaction between alleles conditioning protein and amino acids and how to best proceed with marker-assisted selection (MAS). At the present time, not a single commercial cultivar of soybean with the FAO standard total sulfur containing amino acids has been developed due to the primarily due to the pitfalls described in this paper. With continued improvements in breeding technologies in conjunction with continued elucidation of quantitative traits, it is foreseeable in the future that a soybean cultivar with high yield, protein, and levels of essential sulfur-containing amino acids will be developed.

References

Altenbach S.B., Pearson K.W., Leung F.W., Sun S.S.M. (1987) Cloning and sequence analysis of a cDNA encoding a Brazil nut protein exceptionally rich in methionine. Plant molecular biology 8:239-250.

Altenbach S.B., Pearson K.W., Meeker G., Staraci L.C., Sun S.S.M. (1989) Enhancement of the methionine content of seed proteins by the expression of a chimeric gene encoding a methionine-rich protein in transgenic plants. Plant molecular biology 13:513-522.

Annicchiarico P. (2002) Genotype x environment interactions: challenges and opportunities for plant breeding and cultivar recommendations Food & Agriculture Org.

Bajjalieh N. (2004) Proteins from oilseeds, FAO.

Boerma H., Hussey R., Phillips D., Wood E., Rowan G., Finnerty S. (2000) Registration of'Boggs' soybean. Crop Science 40:294-295.

Brim C. (1966) A modified pedigree system of selection in soybeans. Crop Sci 6:220.

Brummer E., Graef G., Orf J., Wilcox J., Shoemaker R. (1997) Mapping QTL for seed protein and oil content in eight soybean populations. Crop Science 37:370-377.

Burton J., Wilson T. (1998) Registration of 'Prolina'soybean. Crop Science 39:294.

Burton J., Purcell A., Walter Jr W. (1982) Methionine concentration in soybean protein from populations selected for increased percent protein. Crop Sci 22:430-432.

Burton J., Carter Jr T., Bowman D. (2005) Registration of NC-Roy soybean. Crop Sci 45:2654.

Chung J., Babka H., Graef G., Staswick P., Lee D., Cregan P., Shoemaker R., Specht J. (2003) The seed protein, oil, and yield QTL on soybean linkage group I. Crop Science 43:1053.

Churchill G.A., Doerge R.W. (1994) Empirical threshold values for quantitative trait mapping. Genetics 138:963.

Clarke E., Wiseman J. (2000) Developments in plant breeding for improved nutritional quality of soya beans I. Protein and amino acid content. The Journal of Agricultural Science 134:111-124.

Council N.R. (1994) Nutrient Requirements of Poultry, National Academy Press, Washington, DC.

Council N.R. (1998) (Ed.)^(Eds.) Nutrient Requirements of Swine, National Academy Press, Washington, DC. pp. Pages.

Danielsson C. (1949) Seed globulins of the Gramineae and Leguminosae. Biochemical Journal 44:387.

Diers B., Keim P., Fehr W., Shoemaker R. (1992) RFLP analysis of soybean seed protein and oil content. TAG Theoretical and Applied Genetics 83:608-612.

Edwards 3rd H., Douglas M., Parsons C., Baker D. (2000) Protein and energy evaluation of soybean meals processed from genetically modified high-protein soybeans. Poultry Science 79:525.

Falco S.C., Guida, T., Locke M., Mauvais, J., Sanders, C., Ward, R.T., and Webber, P. (1995) Transgenic canola and soybean seeds with increased lysine. Bio/Technology 13:577-582.

Falconer D., Mackay T. (1996) Introduction to quantitative genetics (1996). Essex: Longman.

Friedman M., Brandon D. (2001) Nutritional and Health Benefits of Soy Proteins†. J. Agric. Food Chem 49:1069-1086.

George A.A., De Lumen B.O. (1991) A novel methionine-rich protein in soybean seed: Identification, amino acid composition, and N-terminal sequence. Journal of Agricultural and Food Chemistry 39:224-227.

Grabau L.J., Blevins D.G., Minor H.C. (1986) Stem infusions enhanced methionine content of soybean storage protein. Plant physiology 82:1013.

Grant D.w.s.o., Imsande M., Shoemaker R. (2002) SoyBase, the USDA-ARS soybean genome database. Published online by USDA-ARS.

Hyten D., Choi I., Song Q., Specht J., Carter Jr T., Shoemaker R., Hwang E., Matukumalli L., Cregan P. (2010a) A High Density Integrated Genetic Linkage Map of Soybean and the Development of a 1536 Universal Soy Linkage Panel for Quantitative Trait Locus Mapping.

Hyten D., Song Q., Choi I., Yoon M., Specht J., Matukumalli L., Nelson R., Shoemaker R., Young N., Cregan P. (2008a) High-throughput genotyping with the GoldenGate assay in the complex genome of soybean. TAG Theoretical and Applied Genetics 116:945-952.

Hyten D.L., Choi I.Y., Song Q., Specht J.E., Carter Jr T.E., Shoemaker R.C., Hwang E.Y., Matukumalli L.K., Cregan P.B. (2010b) A High Density Integrated Genetic Linkage Map of Soybean and the Development of a 1536 Universal Soy Linkage Panel for Quantitative Trait Locus Mapping. Crop Science 50:960-968.

Hyten D.L., Song Q., Choi I.Y., Yoon M.S., Specht J.E., Matukumalli L.K., Nelson R.L., Shoemaker R.C., Young N.D., Cregan P.B. (2008b) High-throughput genotyping with the GoldenGate assay in the complex genome of soybean. TAG Theoretical and Applied Genetics:1-8.

Imsande J. (2001) Selection of soybean mutants with increased concentrations of seed methionine and cysteine. Crop Science 41:510-515.

Imsande J., Schmidt J.M. (1998) Effect of N source during soybean pod filling on nitrogen and sulfur assimilation and remobilization. Plant and Soil 202:41-47.

Jansen R.C., Stam P. (1994) High resolution of quantitative traits into multiple loci via interval mapping. Genetics 136:1447.

Kerley M., Allee G. (2003) Modifications in soybean seed composition to enhance animal feed use and value: Moving from a dietary ingredient to a functional dietary component.

Kim S.-D., E.-H. Hong, Y.-H. Kim, S.-H. Lee, Y.-K. Seong, K.-Y. Park, Y.-H. Lee, Y.-H. Hwang, E.-H. Park, H.-S. Kim, Y.-H. Ryu, R.-K. Park, and Y.-S. Kim. (1996) A new high protein and good seed quality soybean variety "Danbaegkong". RDA Journal of Agricultural Science- Upland and Industrial Crops 38:228-232.

Kim W.S., and H.B. Krishnan. (2004) Expression of an 11 kDa methionine-rich delta-zein in transgenic soybean results in the formation of two types of novel protein bodies in transitional cells situated between the vascular tissue and storage parenchyma cells. . J.Plant Biotech. 2:199-210.

Kortt A., Caldwell J., Lilley G., Higgins T. (1991) Amino acid and cDNA sequences of a methionine-rich 2S protein from sunflower seed (Helianthus annuus L.). European journal of biochemistry/FEBS 195:329.

Kosambi D. (1944) The estimation of map distances from recombination values. Ann. Eugen 12:172-175.

Krishnan H. (2008) Improving the Sulfur-Containing Amino Acids of Soybean to Enhance its Nutritional Value in Animal Feed. Sulfur: A Missing Link Between Soils, Crops, and Nutrition:235.

Krober O. (1956) Nutritive Quality of Crops, Methionine Content of Soybeans as Influenced by Location and Season. Journal of Agricultural and Food Chemistry 4:254-257.

Lorieux M. (2007) MapDisto, A Free User-Friendly Program For Computing Genetic Maps.

Mansur L., Lark K., Kross H., Oliveira A. (1993) Interval mapping of quantitative trait loci for reproductive, morphological, and seed traits of soybean (Glycine max L.). TAG Theoretical and Applied Genetics 86:907-913.

Martin O.C. (2006) Two-and three-locus tests for linkage analysis using recombinant inbred lines. Genetics 173:451.

Nakasathien S., Israel D., Wilson R., Kwanyuen P. (2000) Regulation of seed protein concentration in soybean by supra-optimal nitrogen supply.

Nichols D., Glover K., Carlson S., Specht J., Diers B. (2006) Fine mapping of a seed protein QTL on soybean linkage group I and its correlated effects on agronomic traits. Crop Science 46:834.

Nielsen N., Jung R., Nam Y.W., Beaman T., Oliveira L., Bassüner R. (1995) Synthesis and assembly of 11S globulins. Journal of plant physiology 145:641-647.

Nielsen N.C., Dickinson C.D., Cho T.J., Thanh V.H., Scallon B.J., Fischer R.L., Sims T.L., Drews G.N., Goldberg R.B. (1989) Characterization of the glycinin gene family in soybean. The Plant Cell Online 1:313.

Nielsen N.C., M.P. Scott, W.J.P Lago. (1990) Assembly properties of modified subunit family., in: R. H. a. B. Larkins (Ed.), NATO Advanced Study Institute on Plant Molecular Biology, Plenum Press, New York, Schloss Elmau, Germany. pp. p.635-640.

Nyquist W., Baker R. (1991) Estimation of heritability and prediction of selection response in plant populations. Critical Reviews in Plant Sciences 10:235-322.

Paek N., Imsande J., Shoemaker R., Shibles R. (1997) Nutritional control of soybean seed storage protein. Crop Science 37:498-503.

Paek N., Sexton P., Naeve S., Shibles R. (2000) Differential accumulation of soybean seed storage protein subunits in response to sulfur and nitrogen nutritional sources. PLANT PRODUCTION SCIENCE-TOKYO- 3:266-272.

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

Panthee D., Kwanyuen P., Sams C., West D., Saxton A., Pantalone V. (2004) Quantitative trait loci for -conglycinin (7S) and glycinin (11S) fractions of soybean storage protein. Journal of the American Oil Chemists' Society 81:1005-1012.

Panthee D., Pantalone V., Sams C., Saxton A., West D., Orf J., Killam A. (2006b) Quantitative trait loci controlling sulfur containing amino acids, methionine and cysteine, in soybean seeds. TAG Theoretical and Applied Genetics 112:546-553.

Rajcan I., Hou G., Weir A. (2005) Advances in breeding of seed-quality traits in soybean. Journal of Crop Improvement 14:145-174.

Rieseberg L.H., Archer M.A., Wayne R.K. (1999) Transgressive segregation, adaptation and speciation. Heredity 83:363-372.

Sallstrom J.R. (2002) Better Bean Initiative (BBI)-A tool to enhance competitiveness for the U.S. soybean producer, Bienn. Conf. Cell. Mol. Biol. Soybean, Urbana-Champaign.

SAS. (2003) SAS/STAT User's Guide, Version 9.1.

Sebolt A., Shoemaker R., Diers B. (2000) Analysis of a quantitative trait locus allele from wild soybean that increases seed protein concentration in soybean. Crop Science 40:1438.

Serretti C., Schapaugh Jr W., Leffel R. (1994) Amino acid profile of high seed protein soybean. Crop Science 34:207.

Sexton P., Naeve S., Paek N., Shibles R. (1998) Sulfur availability, cotyledon nitrogen: sulfur ratio, and relative abundance of seed storage proteins of soybean. Crop Science 38:983-986.

Sexton P., Paek N., Naeve S., Shibles R. (2002) Sulfur metabolism and protein quality of soybean. Journal of Crop Production 5:285-308.

Shen B., Li C., Tarczynski M.C. (2002) High free methionine and decreased lignin content result from a mutation in the Arabidopsis S adenosyl L methionine synthetase 3 gene. The Plant Journal 29:371-380.

Shewry P., Napier J., Tatham A. (1995) Seed storage proteins: structures and biosynthesis. The Plant Cell Online 7:945.

Townsend J., Thomas L. (1994) Factors which influence the Agrobacterium-mediated transformation of soybean. J. Cell Biochem. Suppl. 18A 78.

Wang S., Basten C., Zeng Z. (2007) Windows QTL Cartographer 2.5. Department of Statistics, North Carolina State University, Raleigh, NC.

Warrington C.V., J.H. Orf, A.S. Kilam, N. Bajjalieh, and H.R. Boerma. (2011) Resource Allocation for Seed Protein and Several Amino Acids in Soybean, Institute of Plant Breeding, Genetics, and Genomics, University of Georgia, Athens.

Weber C., Fehr W. (1970) Registration of Provar soybeans. Crop Sci 10:728.

Wilcox J., Shibles R. (2001) Interrelationships among seed quality attributes in soybean. Crop Science 41:11.

Wilson R. (2004) Seed composition. Soybeans: improvement, production, and uses 3:621–677.

Yaklich R. (2001) [beta]-Conglycinin and Glycinin in High-Protein Soybean Seeds. J. Agric. Food Chem 49:729-735.

Yaklich R., Helm R., Cockrell G., Herman E. (1999) Analysis of the distribution of the major soybean seed allergens in a core collection of Glycine max accessions. Crop Science 39:1444.

Yates J.L. (2006) Agronomic performance and amino acid composition of near-isogenic soybean lines differing for an allele at a major protein QTL, Dept. of Crop and Soil Sciences, University of Georgia, Athens, GA. pp. 155-191.

Table 2.1. Protein and amino acid means for Danbaekkong, three check cultivars, and the means and ranges for 140 Benning × Danbaekkong RILs in five environments.

	Parent means		F_5-derived RILs			
Trait	Dan†	Checks‡	Mean	Min	Max	LSD (0.05)
Protein (g kg^{-1})	510	433	468	425	507	13
Lys/cp (g kg^{-1})	61.2	64.8	63.3	61.2	65.0	1.1
Thr/cp (g kg^{-1})	34.4	37.9	36.1	33.8	38.0	0.7
Met/cp (g kg^{-1})	14.2	14.4	14.2	13.8	14.7	0.4
Cys/cp (g kg^{-1})	15.7	15.3	15.3	14.7	16.2	0.4
Met+Cys/cp (g kg^{-1})	29.9	29.7	29.6	28.5	30.8	0.7

†Danbaekkong

‡Mean of three checks (NCRoy, AG6202, BoggsRR)

Table 2.2. Phenotypic (in bold font) and genetic correlation coefficients among crude protein and amino acids for the Benning × Danbaekkong RIL population.

	Crude Protein	Lys/cp	Thr/cp	Met/cp	Cys/cp	Met+Cys/cp
Crude Protein		**-0.63****	**-0.85****	**-0.19***	**-0.16**	**-0.20****
Lys/cp	-0.82		**0.78****	**0.70****	**0.41***	**0.63****
Thr/cp	-0.91	0.92		**0.44****	**0.41****	**0.47****
Met/cp	-0.36	0.47	0.57		**0.57****	**0.88****
Cys/cp	-0.06	0.17	0.36	0.76		**0.89****
Met+Cys/cp	-0.21	0.32	0.48	0.92	0.95	

*, ** significant at the $P \leq 0.01$ and $P \leq 0.001$ level of significance

Table 2.3. DNA markers associated with crude protein (g kg^{-1}) using single factor analysis of variance (SF-ANOVA; P = 0.01) and composite interval mapping (CIM) for the mean of 140 RILs grown in five environments.

		SF-ANOVA		CIM	
Chromosome (LG)	Marker	2a†	R^2	LOD	R^2
		g kg^{-1}	%	score	%
6 (C2)	BARC-042781	8.8	6		
7 (M)	Satt336	9.0	6		
10 (O)	Satt478	7.5	4		
13 (F)	Satt114	-7.7	4		
14 (B2)	BARC-018353	8.3	5	3.8	5
15 (E)	BARC-027786	8.3	5	4.4	10
17 (D2)	BARC-019505	5.6	2‡	5.1	9
20 (I)	BARC-061899	28.2	53	29.3	55

† 2a, the difference in crude protein content at a marker locus homozygous for Danbaekkong vs. homozygous for Benning. A positive value indicates the allele for increased protein is inherited from Danbaekkong.

‡ significant at P = 0.05

Table 2.4. DNA markers associated with Lys (Lys) as % crude protein (g kg^{-1}) using single factor analysis of variance (SF-ANOVA; P = 0.01) and composite interval mapping (CIM) for the mean of 140 RILs grown in five environments.

			SF-ANOVA		CIM	
Trait	Chromosome (LG)	Marker	2a†	R^2	LOD	R^2
			g kg^{-1}	%	score	%
Lys/cp	5 (C1)	BARC-024445	-0.49	7		
	8 (A2)	BARC-055265	-0.31	4‡	3.4	6
	15 (E)	Satt231	-0.35	4		
	20 (I)	BARC-061899	-1.20	47	23.2	48
Thr/cp	1 (D1a)	BARC-035219	-0.27	3‡	3.3	6
	4 (C1)	BARC-024445	-0.39	4		
	7 (M)	Satt336	-0.34	5		
	9 (K)	BARC-048619	-0.28	3	3.9	5
	10 (O)	Satt478	-0.40	6		
	17 (D2)	Satt256	-0.10	ns	3.7	6
	20 (I)	BARC-061899	-1.30	51	26.1	53
Met/cp	6 (C2)	BARC-055889	-0.06	3‡	3.6	8
	9 (K)	BARC-042449	-0.11	8	4.1	9
	10 (O)	Satt592	-0.14	11	5.9	14
	20 (I)	BARC-020713	-0.16	12	7.7	20
Cys/cp	6 (C2)	BARC-048543	0.11	5		
	10 (O)	Satt592	-0.20	13	3.5	10
	14 (B2)	BARC-016831	0.11	5		
	20 (I)	BARC-020713	-0.14	6		

Table 2.4 continued.

			SF-ANOVA		CIM	
Trait	Chromosome (LG)	Marker	2a†	R^2	LOD	R^2
Met+Cys/cp	6 (C2)	BARC-047715	0.17	4		
	10 (O)	Satt592	-0.34	15	4.3	11
	14 (B2)	BARC-016831	0.16	4		
	18 (G)	BARC-039397	0.22	5		
	20 (I)	BARC-020713	-0.30	10	4.3	12

† 2a, the difference in crude protein content at a marker locus homozygous for Danbaekkong vs. homozygous for Benning. A positive value indicates the allele for the increased trait value is inherited from Danbaekkong.

‡ significant at P = 0.05

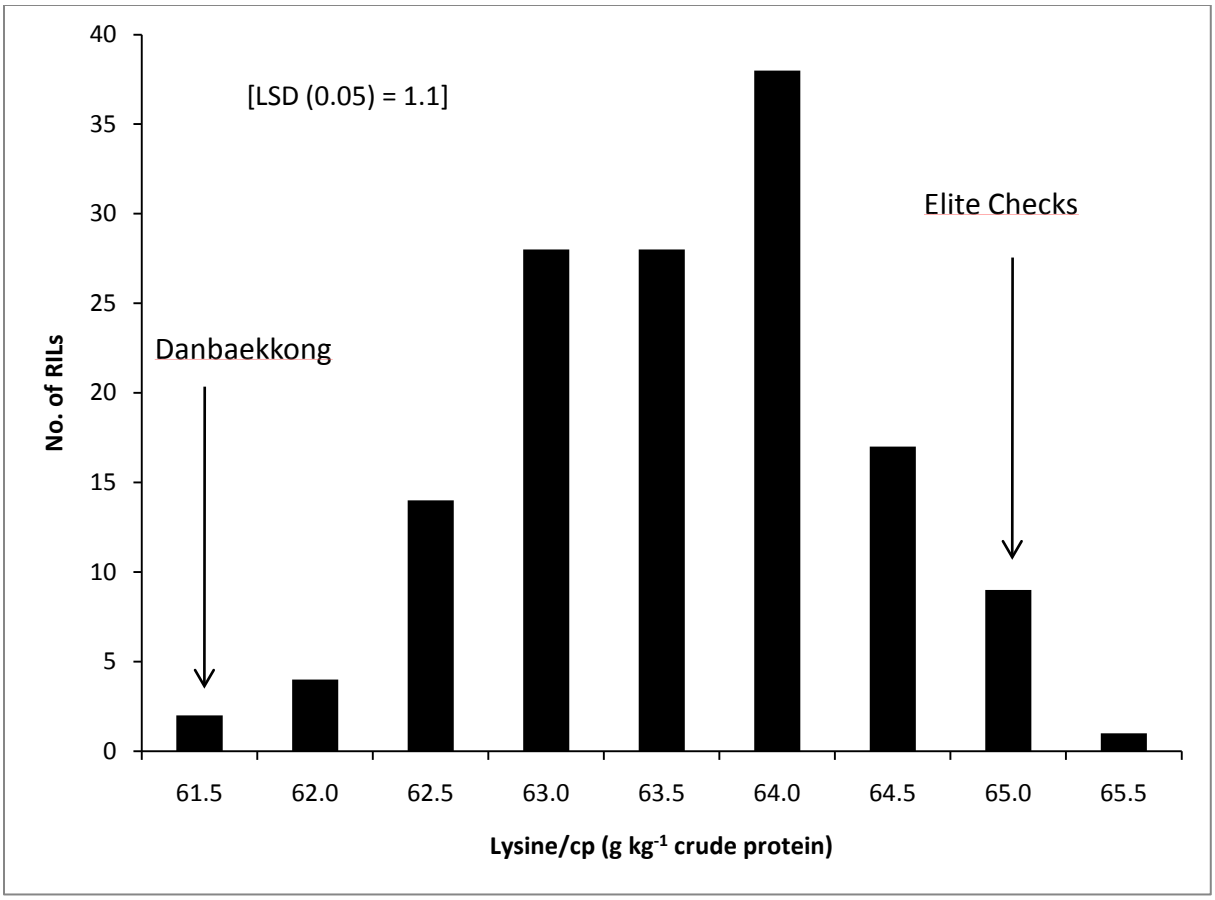


Figure 2.1. Distribution of Lys/cp (g kg^{-1} crude protein) in the Benning × Danbaekkong RIL population.

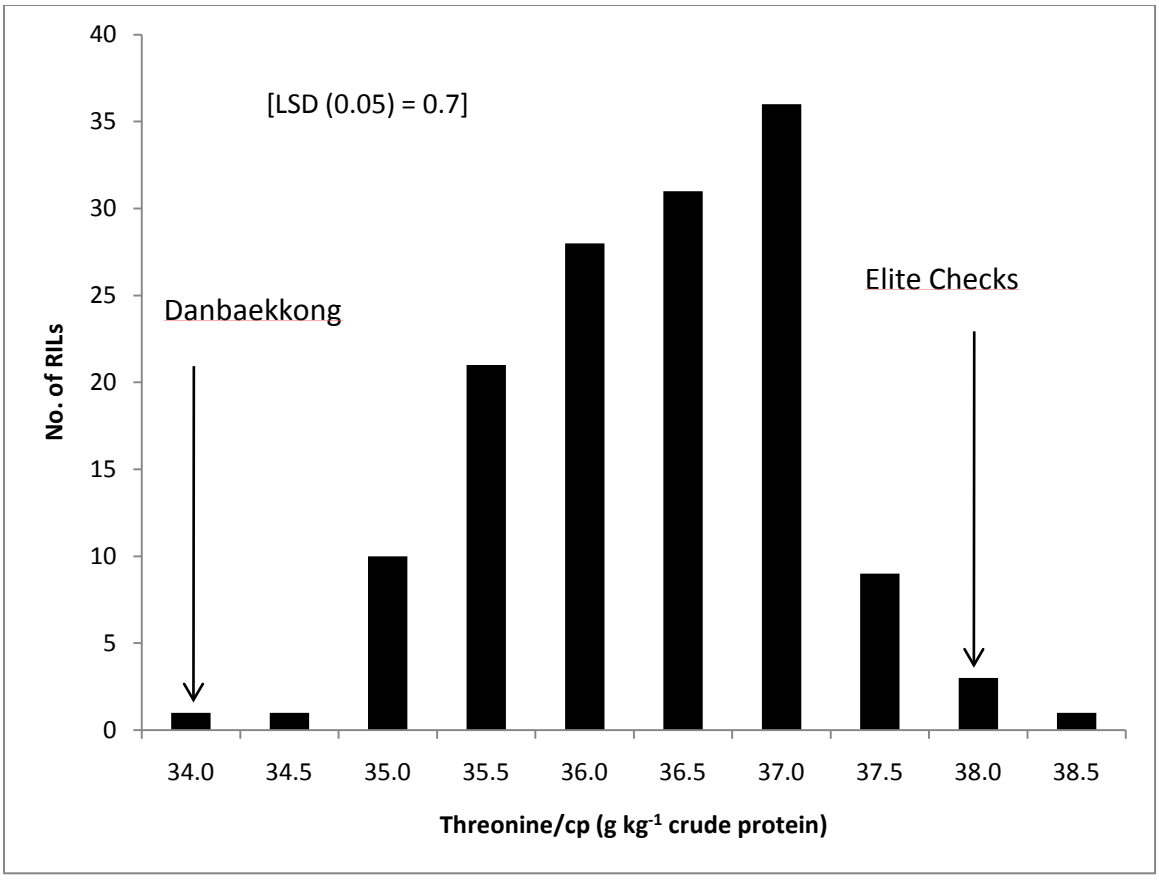


Figure 2.2. Distribution of Thr/cp (g kg^{-1} crude protein) in the Benning × Danbaekkong RIL population.

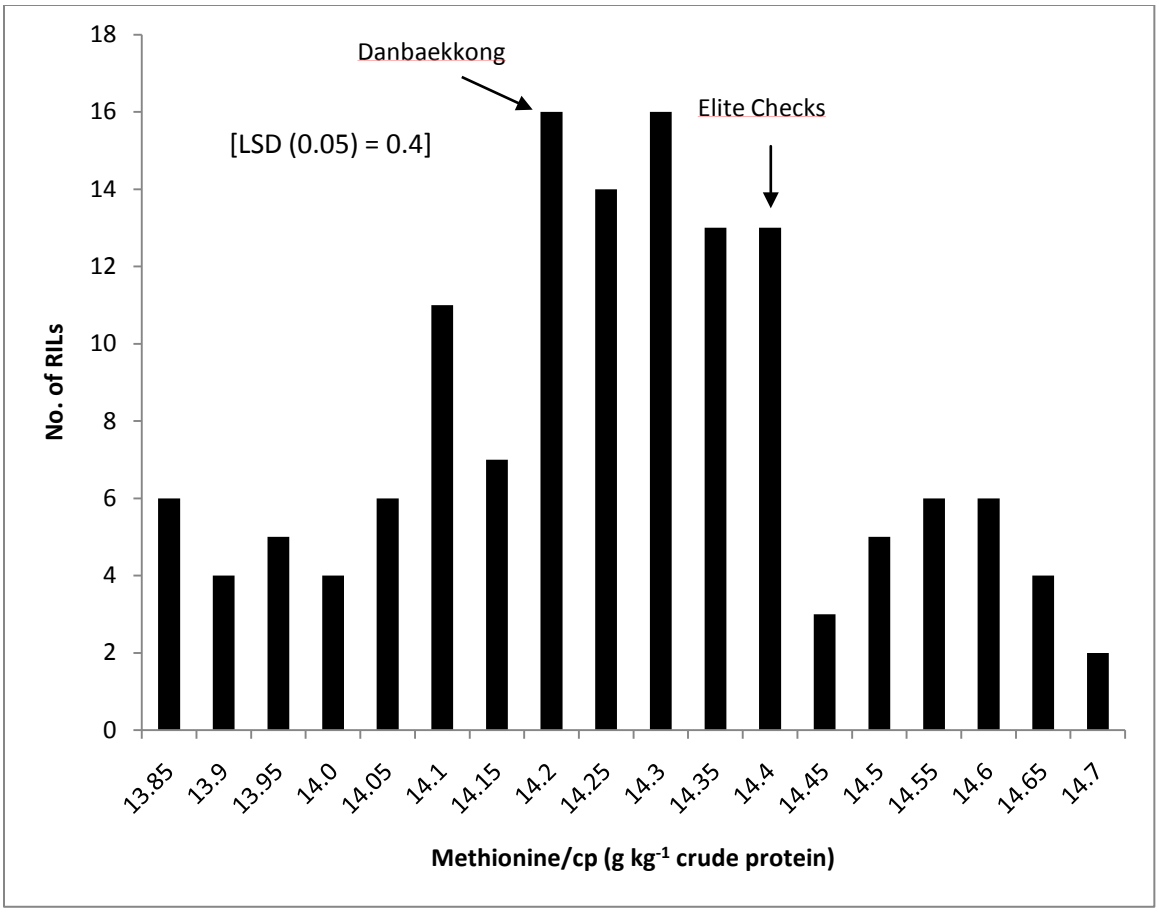


Figure 2.3. Distribution of Met/cp (g kg^{-1} crude protein) in the Benning × Danbaekkong RIL population.

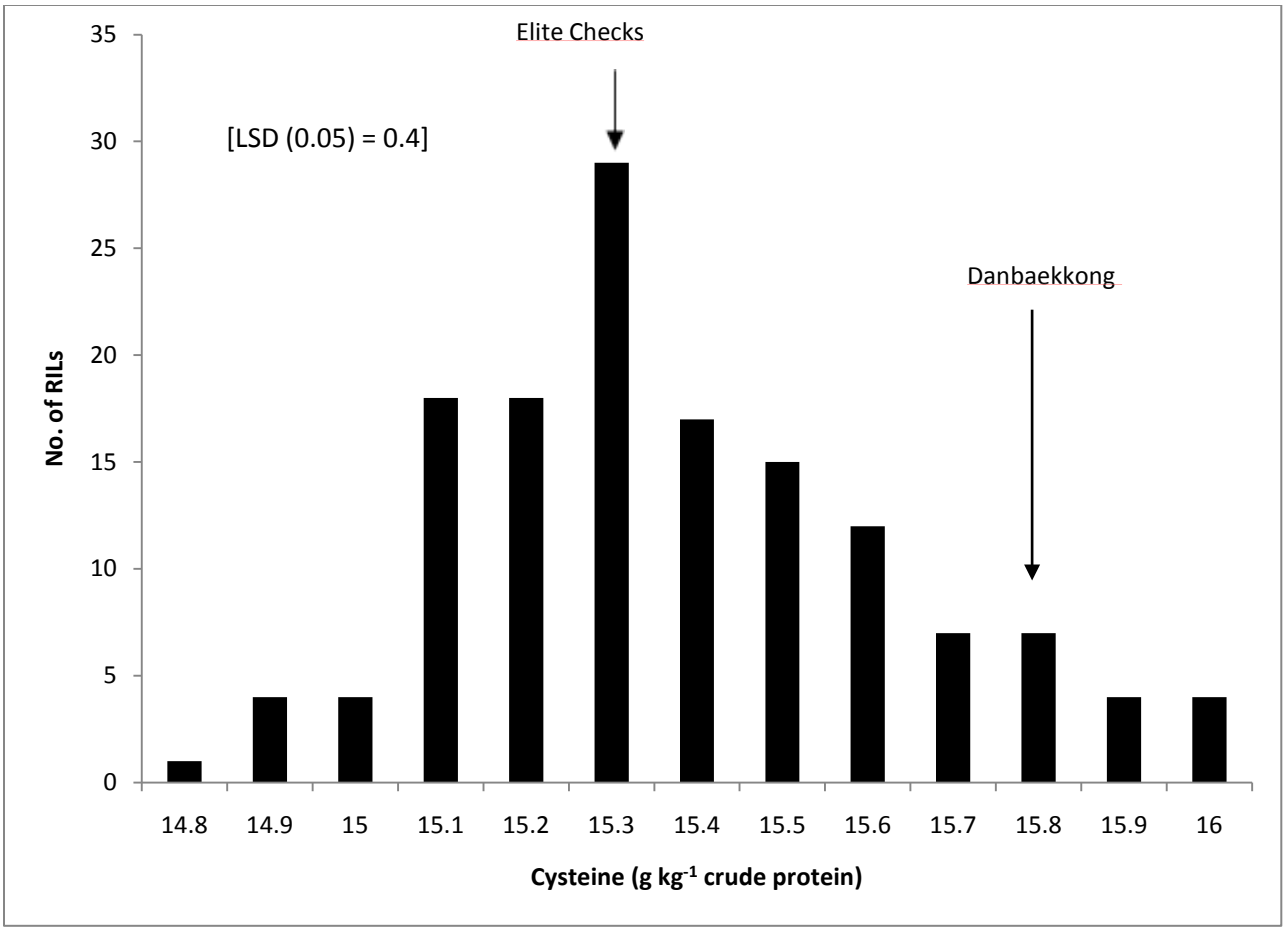


Figure 2.4. Distribution of Cys/cp (g kg^{-1} crude protein) in the Benning × Danbaekkong RIL population.

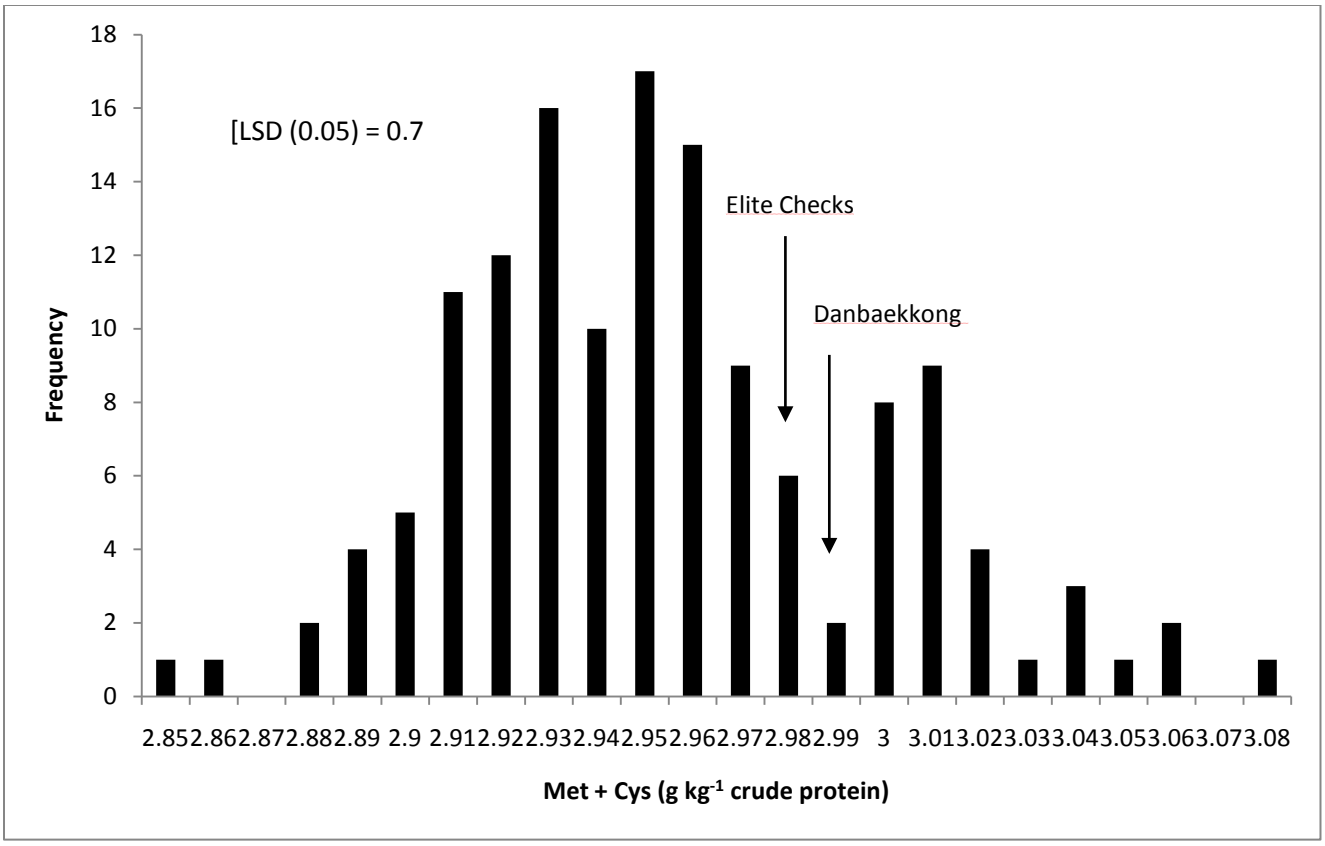


Figure 2.5. Distribution of Met+Cys/cp (g kg^{-1} crude protein) in the Benning × Danbaekkong RIL population.

Figure 2.6 Comparison between Benning × Danbaekkong population linkage map (shown on right) and the integrated genetic linkage map (Consensus Map 4.0 (Hyten et al, 2010a) (shown on left). Bars shown are connecting identical marker loci.

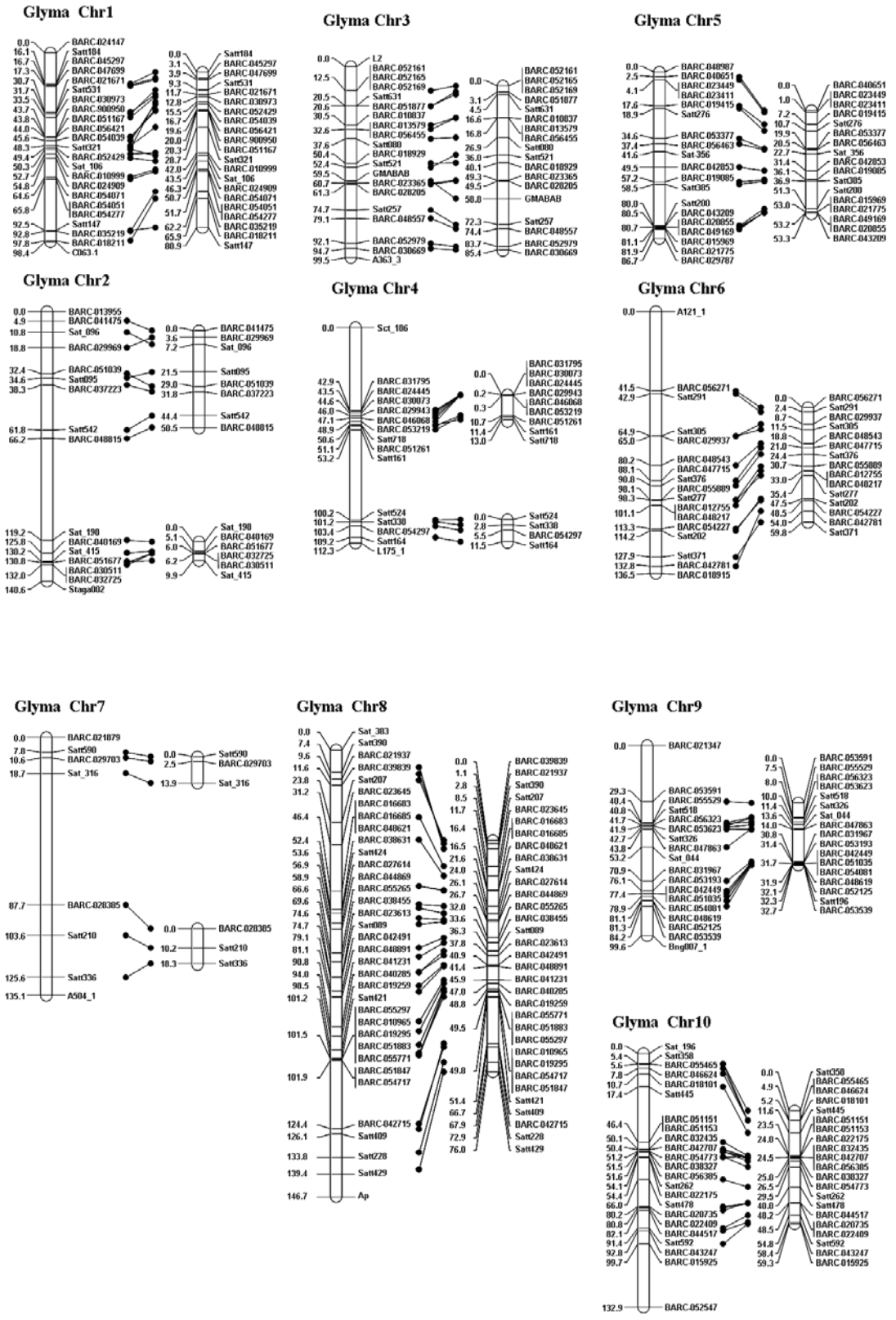


Figure 2.6 continued.

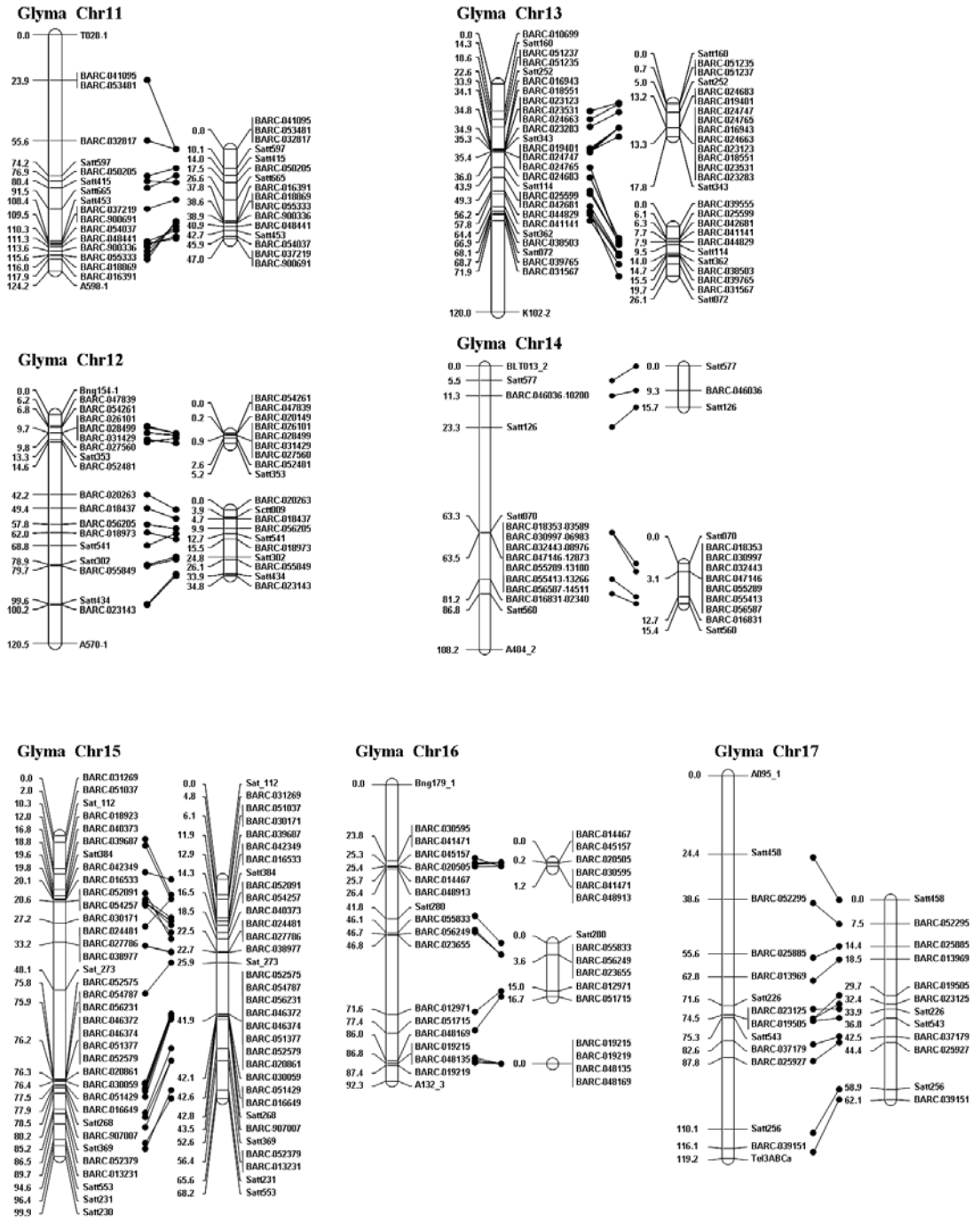


Figure 2.6. continued.

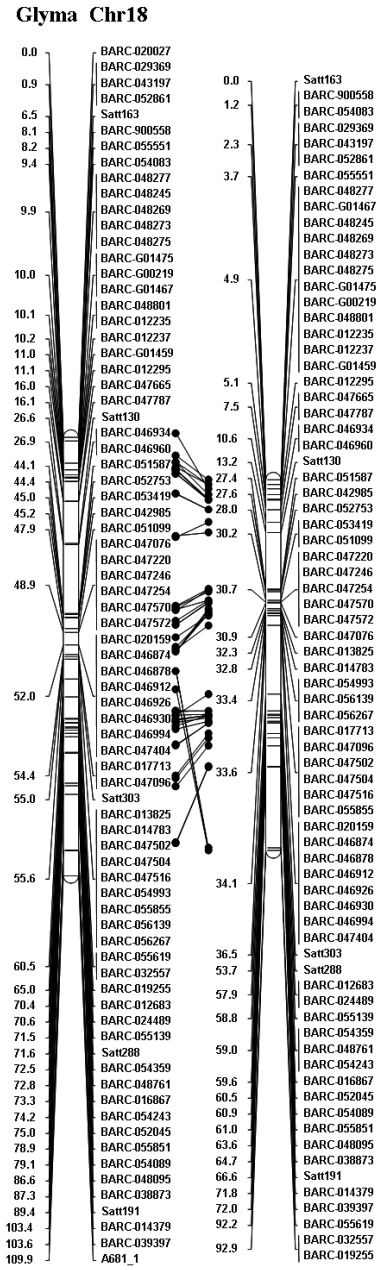


Figure 2.6. continued.

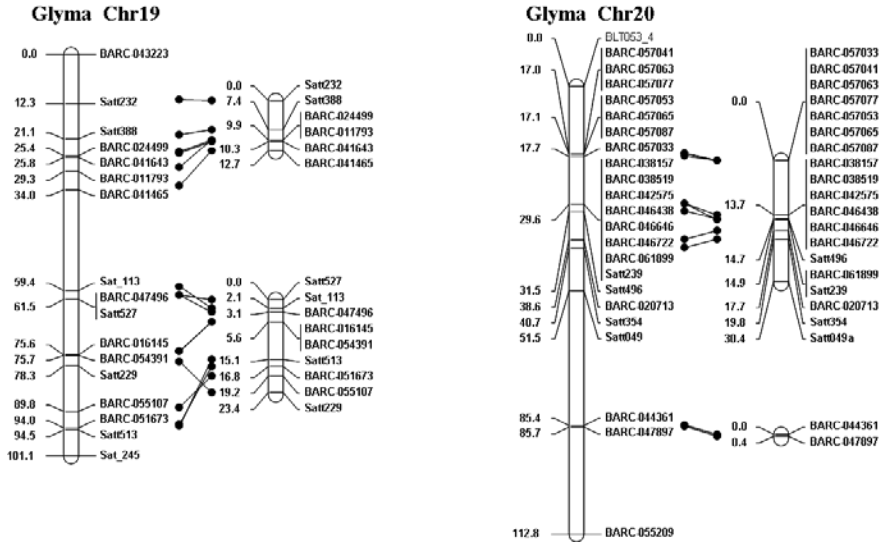


Figure 2.7. QTL likelihood plots from composite interval mapping (CIM) for crude protein and amino acid QTL using 140 recombinant inbred lines (RILs) from the Benning × Danbaekkong population. For each chromosome (Chr), the permutation-derived (n = 1000 per trait) LOD score significance criteria are indicated by a vertical dotted line at the threshold level of 3.1 for each trait. Crude protein (————); Lys (— · · ·); Thr/cp (– – –); Met/cp (·····); Cys/cp (— —); Met+Cys/cp (— · —).

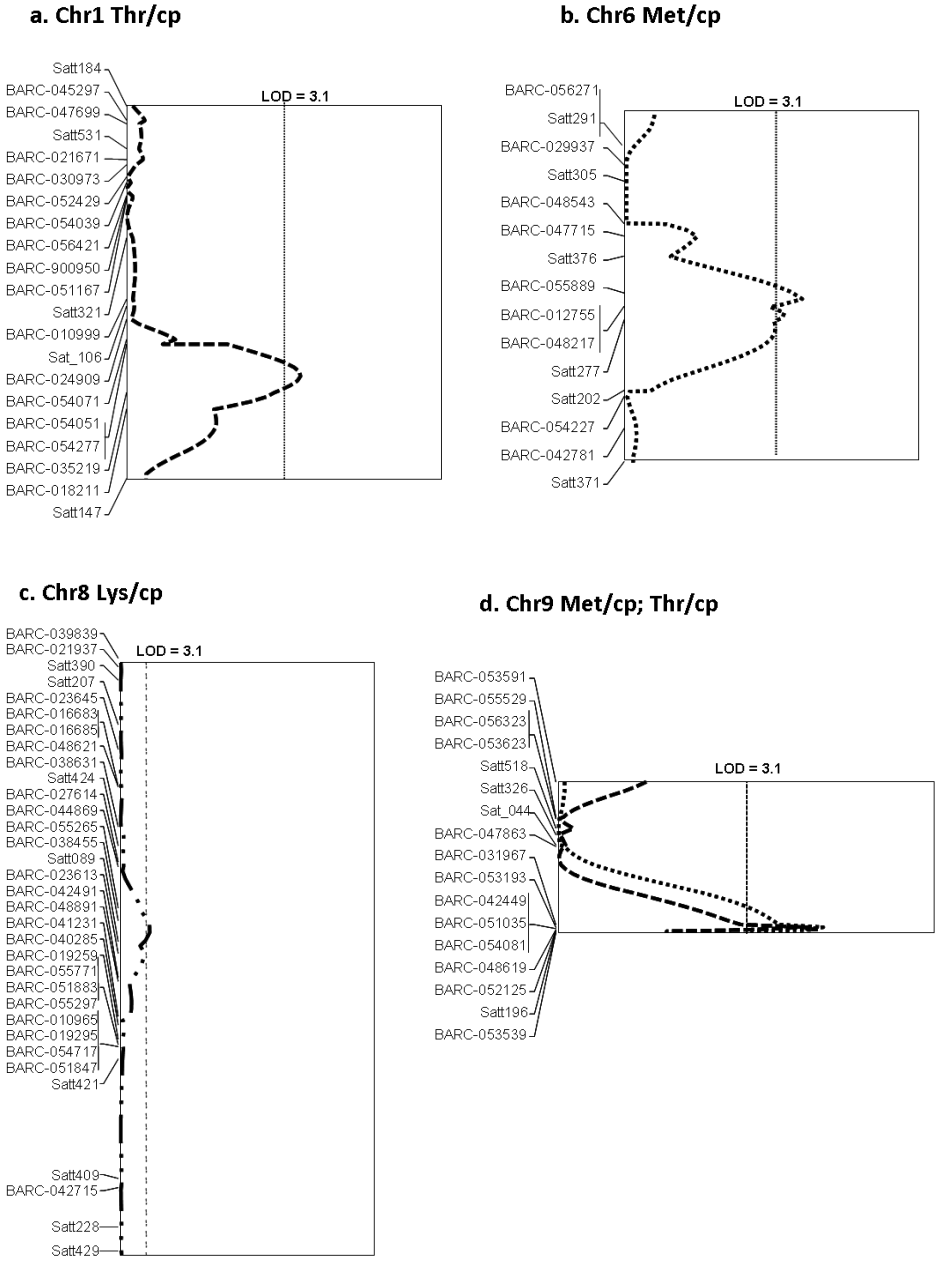
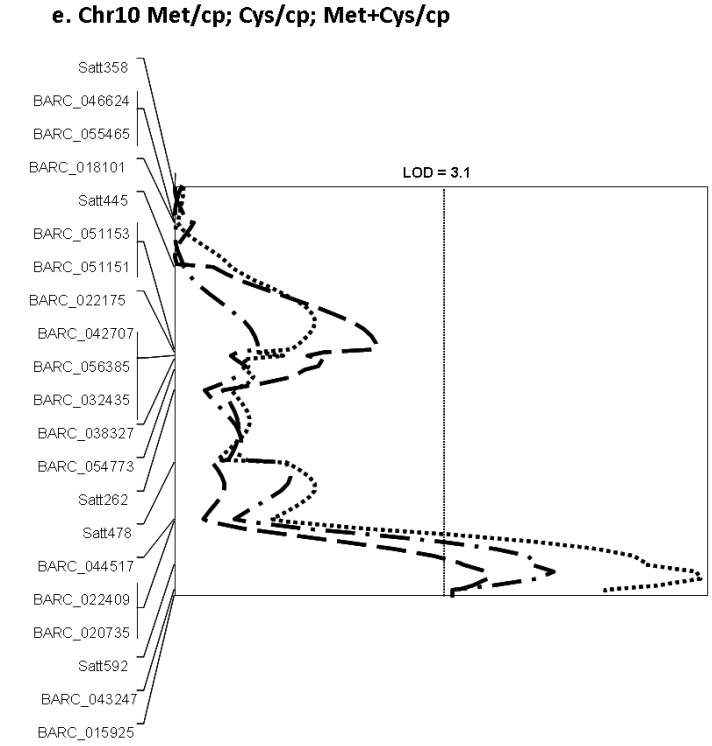
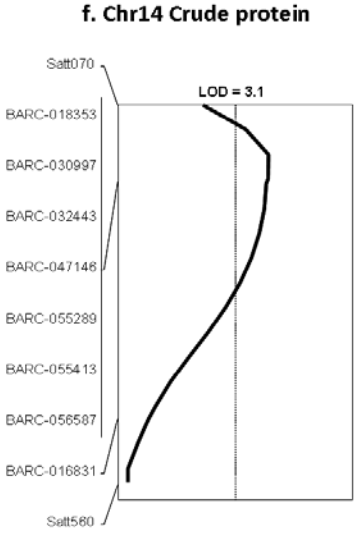


Figure 2.7. continued.





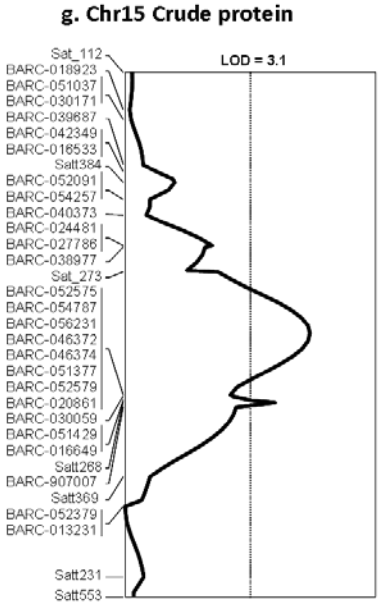
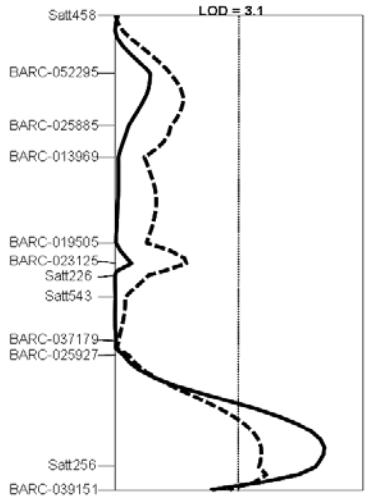
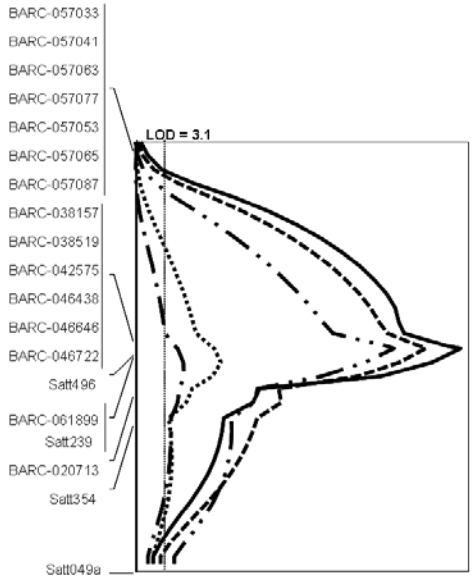


Figure 2.7.continued.

h. Chr17 Crude protein; Thr/cp



i. Chr20 Crude protein; Lys/cp; Thr/cp; Met/cp; Met+Cys/cp



CHAPTER 4

RESOURCE ALLOCATION FOR SEED PROTEIN AND SEVERAL AMINO ACIDS IN SOYBEAN[1]

[1] Warrington, C.V., J.H. Orf, A.S. Killam, N. Bajjalieh, and H.R. Boerma. To be submitted to *Crop Science*

Abstract

Soybean [*Glycine max* (L.) Merr.] is the world's leading provider of oil and high-quality protein meal. An understanding of genotype × environment (G×E) interactions associated with crude protein (cp), lysine (Lys), threonine (Thr), methionine (Met), and cysteine (Cys) would aid breeders in their selection efforts. The objectives of this study were to determine the importance of G×E interactions for protein and amino acid content, to assess the optimum number of replications and environments necessary to provide a given level of discrimination among genotypes for crude protein and amino acids, and to evaluate the association of seed yield, maturity, and other agronomic traits with amino acid content. To meet these objectives, 140 F_5-derived recombinant inbred lines (RILs) were developed from a cross of 'Benning' × 'Danbaekkong' and were grown in five field environments across the southern USA. The effects of genotype and genotype × environment interaction were significant for crude protein, Lys/cp, Thr/cp, Met/cp, Cys/cp, and Met+Cys/cp ($P \leq 0.001$). The genotypic variance component for crude protein was roughly seven times larger than the G×E variance component for this trait. These two components of variance were found to be of similar magnitude for Lys/cp and Cys/cp. The G×E component of variance was slightly higher than the genotypic component for Thr/cp, Met/cp and Met+Cys/cp. It was determined that the combination of five replications and two environments used in our study could detect a difference of 2.5% between two RIL means for Lys/cp and Thr/cp. An increased number of plots (environment/replication combinations) would be necessary to detect a 2.5% difference or less between two RIL means for crude protein, Met/cp, Cys/cp, and Met+Cys/cp. Significant correlations were detected between yield and crude protein (r = -0.50), Lys/cp (r = 0.49), and Thr/cp (r = 0.49). Crude protein and amino acid associations with maturity, seed weight, plant height, and lodging were negligible.

Introduction

Soybean [*Glycine max* L. (Merr.)] is the world's leading provider of oil and high-quality protein meal. The amino acid constituents of the protein are the major determinate in how efficient the meal is in providing growing poultry, swine, and other livestock with the necessary nutrition for maximum health and growth. The United Soybean Board's Better Bean Initiative (BBI) included as one of its major research goals in 2002 that soybean breeders strive towards developing soybean cultivars in the USA with improved amino acid quality. One of the primary traits targeted by the BBI is increased methionine and cysteine while increased levels of lysine and threonine represent secondary goals of the initiative. It is imperative for breeders that develop soybean cultivars with modified amino acid profiles as well as growers that produce the improved soybeans seeds are aware of genotype × environment (G×E) interaction as it relates to the amino acid profile of soybean and their overall agronomic performance.

As quantitative traits, protein and amino acid concentrations are controlled by a number of genes and are affected by the environment to a degree (East, 1916). Genotype × environment (G×E) interaction refers to differential genotypic responses in different environments, which reduces the association between genotype and phenotype. Factors such as temperature, soil moisture, soil type, and fertility level which fluctuate among environments, either location or year, potentially contribute to inconsistent genotypic responses. When G×E interaction is present, the effects of genotypes and environments are statistically nonadditive, implying that the differences between genotypes are dependent upon the specific environment in which they are grown (Hühn, 1996; Yue et al., 1997). The knowledge of G×E interactions and stability of genotypes across environments is vital for the development of an efficient and effective breeding strategy to modify the amino acid profile in soybean. For plant breeders, dealing with G×E

interaction may be viewed in a number of ways. Breeders may deal with large G×E interactions by developing breeding programs and selecting genotypes within more homogeneous sub-regions of a larger area. This method of dealing with G×E interaction is costly and does not overcome the influence of year × genotype interactions over the entire target environment area (Scapim et al., 2000). The presence of G×E interaction requires that breeders test genotypes in the appropriate environmental conditions likely to be encountered in the target environments where the genotypes are to be grown. Thus, breeders aim to develop stable cultivars with good performance over a range of environmental conditions (Weber et al., 1996). It is important to elucidate the degree to which G×E interactions influence the expression of a prospective 'value-added' change in seed composition (Wilson, 2004). There is limited information concerning the G × E interactions associated with amino acid content in soybean, though protein and other value-added traits have been investigated previously. Breeding programs have been fairly successful in increasing protein (Brim and Burton, 1979; Burton and Wilson, 1998; Miller, 1979; Weber and Fehr, 1970; Wilcox, 1998; Wilcox and Cavins, 1995), but environmental variation can make selection for this trait difficult, even when heritability is high (Brummer et al., 1997). Fehr (2003) determined G×E interactions were not significant in regard to the protein components β-conglycinin and glycinin, an indication that breeders should have success in breeding for soybean lines with varied levels of protein components. Temperature during the growing season has been shown to have an affect on protein concentrations in soybean (Wolf et al., 1982), although it is evident that there is much variability in plant response to increasing or decreasing temperatures (Dornbos and Mullen, 1992; Gibson and Mullen, 1996; Sato and Ikeda, 1979). The causal basis for the seed constituent response is unknown, however it may be that the effectiveness of the metabolic machinery is affected by temperature, thus resulting in G×E

interactions (Wilson, 2004). Studies of the effects of different environments on soybean fatty acids showed that increased temperature was a determinant in reducing linolenic and increasing oleic acid concentrations in soybean seed (Cherry et al., 1985). A study by Primomo (2002) found that the genotype × year interaction was significant for all fatty acids tested, but that only oleic, linoleic, and linolenic acids had significant genotype × location and genotype × year × location effects.

An integral component in establishment of markets for soybean cultivars with value-added traits is the determination of how traits respond to changes in environment. The most pragmatic and cost-effective situation would be one in which cultivars express wide adaptation over a range of growing environments. Since yield testing is a costly and time-consuming endeavor, it is necessary that breeders determine the optimum combination of replicates and environments to sufficiently measure the genetic value of a genotype. Increased replications at each location function to improve the precision of measurements and the power to detect differences between genotypes within the location. Conversely, the presence of G×E interaction between the test environment and the breeder's base population of environments negates the effectiveness of more replicates at a single location. Thus, more precision would be gained in discriminating between genotypes by adding additional testing environments as opposed to replications (Bernardo, 2002). The addition of environments is associated with increased resource input including labor, land, and supplies (Kang, 1997). The best location or combination of locations should provide a measure of the relative potential of genotypes over the target population of environments and maximize genetic variation, and in turn, response to selection (Allen and Rasmusson, 1978). This response to selection is highly associated with trait heritability. The impact of heritability estimates for plant breeders is two-fold. First, heritability

estimates provide a measure of the relative ease with which traits can be selected. Second, the relative change in population mean due to selection is a function of heritability, thus heritability estimates are important in predicting population improvement (Hanson, 1963). With that being said, there appears to be a trade-off between precision and resource allocation.

It is extremely important for breeders to be cognizant of the influence of G×E interaction on seed composition and seed quality traits in soybean and its role in how breeding programs proceed. Although there have been a few studies concerned with G×E interaction in regard to protein, none have been published which have investigated this phenomena for soybean amino acids. The objectives of this study were to: (i) determine the importance of G×E interactions for protein and amino acid content, and (ii) to assess the effect of G×E interaction on selection of genotypes as determined by the optimum number of replications and environments necessary to provide a given level of discrimination among genotypes for crude protein and amino acids, and (iii) evaluate the association of seed yield, maturity, and other agronomic traits with amino acid content.

Materials and Methods

Plant Material

A population of 150 F_5-derived recombinant-inbred lines (RILs) was developed from a cross of Benning (PI595645) (Boerma et al., 1997) × Danbaekkong (PI619083) (Kim, 1996). The parents were chosen based on their disparate protein levels, with Benning averaging approximately 42% and Danbaekkong at 51% on a dry-weight basis. Benning is a high-yielding maturity group VII cultivar adapted to the southeastern USA and Danbaekkong is a South Korean maturity group IV tofu cultivar.

Seeds from the original cross were grown in the greenhouse and seed from individual F_1 plants were bulked. The F_2 plants were grown at the Univ. of Georgia Plant Sciences Farm near Watkinsville, GA. Individual F_2 plants were advanced to the F_5 generation in Athens, GA and at the USDA winter nursery near Isabela, Puerto Rico using a modified single seed descent (Brim, 1966). The F_3 and F_4 generations were grown in Puerto Rico and the F_5 generation was grown at the Univ. of Georgia Plant Sciences Farm. At maturity individual F_5 plants were single-plant threshed to create F_5-derived RILs. Approximately 200 RILs were grown in 2003 and 150 RILs were selected for uniform maturity.

In 2005 and 2006, 150 RILs were planted at the Univ. of Georgia Plant Sciences Farm near Athens, GA and also planted in Bay, AR, Stuttgart, AR, and Kinston, NC in 2006. The 150 RILs were sub-divided into three sets of 50 RILs based on their relative maturity. Danbaekkong and three check cultivars, 'NCRoy', 'AG6202', and 'Boggs-RR' were included in each set. For each set the experimental design was a randomized complete block with two replications. Each set was also randomized within a single replication.

The experiment was planted in Athens in an Appling loamy coarse sand soil type on 19 May 2005 and on a Cecil coarse sandy loam (fine, kaolinitic, thermic Typic Kanhapludults) soil type on 22 May 2006 and were irrigated. The experimental unit in Athens was a 2-row plot that was 7-m long with 76-cm between rows and was seeded with approximately 27 seeds $^{-1}$m row. At maturity, all plots were end-trimmed to a final row length of 3.66 m and the plots were harvested by plot combine. The experiments were planted in a sandy loam soil type in Bay, AR on 13 June 2006 and were irrigated. The experimental unit was a 2-row plot with 76-cm between rows. In Stuttgart, AR, the experiments were planted in a silt loam soil type on 26 May 2006 and were irrigated. The experimental unit was a 2-row plot with 76-cm between rows. The

Kinston, NC experiments were planted on 19 June 2006 in a Portsmouth soil type (loam with ~4% organic matter) (Fine-loamy over sandy or sandy-skeletal, mixed, semiactive, thermic Typic Umbraquults) and were not irrigated. The experimental unit was a 1-row plot harvested from within 3 planted rows spaced 96-cm apart. The plots were end-trimmed to 4.26 m prior harvesting the middle row of each plot.

Data on a plot basis for seed yield, maturity, seed weight, plant height, and lodging were collected. Seed yield was recorded on 130 g kg^{-1} moisture basis in kg ha^{-1}. Maturity was based on the date in which at least 95% of pods were mature, or the R8 stage of development (Fehr et al., 1977). Plant height was measured as the average of three plants from the ground to the terminal node. Lodging scores were based on a rating between 1 and 5, with 1 being erect plants and 5 being prostrate within an entire plot. Seed weight was measured from a 100-seed sample from each plot and reported as mg $seed^{-1}$.

Protein and Amino Acids

Soybean samples were scanned by near infrared (NIR) spectrometry for crude protein (cp), lysine (Lys), threonine (Thr), methionine (Met), and cysteine (Cys) at the Univ.of Minnesota Soybean Breeding Project laboratory. Whole soybean samples were first ground using a Perten LM 3600 grinder and then scanned on a FOSS 6500 NIR Instrument and were reported in g kg^{-1} on a moisture-free basis. The crude protein and amino acid analyses were conducted on 25-g whole seed samples with near-infrared reflectance (NIR). NIR spectra from the FOSS 6500 were predicted using ISIPredict Software version 1.10.2.4842. Each amino acid sample was reported as a proportion of overall crude protein content (g kg^{-1}cp).

Statistical Analyses

A combined analysis of variance was conducted for the protein and amino acid data over the five environments using PROC GLM (SAS, 2003). All effects (environments, replication, sets, and RIL) were considered random in the statistical model. Variance-component heritability estimates were calculated on an entry-mean basis (Nyquist and Baker, 1991) using the following equation:

$$H^2 = \frac{\sigma_g^2}{\sigma_g^2 + (\sigma_g^2/e) + (\sigma^2/re)}$$

where H^2 represents broad-sense heritability, σ^2_g is genotypic variance, σ^2_{ge} is G×E variance, σ^2 is error variance, r is the number of replications, and e is the number of environments. Restricted maximum likelihood (REML) was used to generate variance components for calculating trait heritabilities.

To determine the number of environments (E) required to detect specific differences between treatments with different levels of G×E interaction the following formula which was modified from Mendenhall and Schaeffer (1973) was used:

$$E \geq [2(t_{\alpha/2} + t_{\beta})^2 (\sigma^2/r + \sigma^2_{GE}]/d^2$$

where $t_{\alpha/2}$ is the t-value associated with the significance level of the t-test ($\alpha = 0.05$ for our calculations), σ^2 is the error variance, and *d* is the difference between trait means in g kg^{-1}. The *t*-values are dependent on the degrees of freedom (df) associated with the sample variance. In our calculations, 15 df was specified arbitrarily due to the fact that there are usually 15 or more df in the error terms of most analyses (the df for *t*-values do not seriously affect results). In this equation, r represents the number of replications/environment and σ^2_{GE} is the G×E variance component. The justification for determining the minimum number of replications or environments to detect a specific difference between genotypic means instead of overall

treatment effects has been discussed previously by Boerma et al. (1985), Carter Jr et al. (1983), and Reese et al. (1988).

Results and Discussion

The effect and contribution of each source of variation to crude protein and amino acid concentrations were evaluated through a combined analysis of variance (ANOVA) over five environments (Table 3.1). The G×E interaction effects were significant ($P \leq 0.001$) in the RIL population for crude protein, Lys/cp, Thr/cp, Met/cp, Cys/cp, and Met+Cys/cp. Genotypes were significant ($P \leq 0.0001$) for all traits using the G × E mean squares as the error term for the test of significance. This indicates there were genotypic differences among RILs for crude protein, Lys, Thr, Met, Cys, and Met+Cys even in the presence of significant G×E interaction detected in our study ($P = 0.001$).

The significance of the F-tests is of less importance than the size of the interaction components relative to the size of the genotypic variance, if selection is to be effective for genotypes (Schutz, 1967). Environment was the most important source of variation for all the amino acids while genotype was the most prominent for crude protein. For crude protein and Thr/cp, the genotypic variance component was greater than the G×E component, with the genotypic component roughly seven times greater than that of the G×E component for crude protein. These results are in accordance with those previously reported for protein (Vollmann et al., 2000) and protein components (Fehr et al., 2003). The magnitudes of the G×E interaction variances were between 15 and 37% of the error variances for all the traits measured. Earlier research by Kwon and Torrie (1964) evaluated two soybean populations in the F_3, F_4, and F_5 generations for G×E interaction . The line or genotype × year variance component estimates were larger than either the line × location or the line × location × year variance components for

yield, seed weight, lodging, days to flowering, and percent oil. The genotypic variance component for protein was greater than the interaction components, but less than the error variance. Similarly, Erikson (1982) found that the genotype × location × year variance component was larger than either two-way component (genotype × location and genotype × year), but less than the genotypic variance component.

Due to its prevalence for crude protein and amino acids, the G×E interaction should not be ignored and warrants the testing of genotypes in multiple environments in order to detect and select lines with the desired level of these traits. The optimal allocation of resources based on the ability to detect differences among genotypes for protein and amino acids were assessed based on our computed variance components and given Type I ($\alpha = 0.05$) and Type II ($\beta = 0.5$) error probabilities. Each scenario presented in terms of resource allocation is dependent upon the researcher's desired level of difference in detection among genotypes (ie. 1.25, 2.5, 5.0, or 10.0% of the overall population mean) (Table 3.2). The cost of increasing replications and environments is of great importance in terms of the allocation of resources in a breeding program. In our case, the optimum allocation of resources would be the fewest number of environments and replications that can be used to detect the desired level of difference among genotypes.

Based on our results, increasing the number of environments and replications increases the precision or reduces the value required to detect a difference between two genotypic means. In the testing configuration of five environments and two replications used in the current study, genotypic differences for crude protein could be detected at a level of 5% or greater of the population mean (46.8 g kg^{-1}). Based on the magnitude of difference for crude protein at the 2.5% level, breeders would be required to test between 12 (4 reps × 3 environments) and 16 (2

reps × 8 environments) plots to detect a 11.7 g kg^{-1} difference (Table 3.3). The effect of resource allocation on potential genetic gain can also be interpreted in regard to how heritability is altered. The number of environments and replications plays an important role in the heritability equation, as an increase in either reduces the value of the denominator, thus increasing the heritability value for a trait. Additionally, the amount of genetic variance (numerator) also greatly impacts the heritability value. For crude protein, the heritabilities calculated in our population were reduced from 0.93 to 0.80 as the combination of reps and environments changed from five environments and two reps one environment with two reps. These heritability values are slightly higher than other estimates (Byth et al., 1969; Fehr, 1968; Kwon and Torrie, 1964; Shannon et al., 1972; Smith and Weber, 1968) and reflect a large amount of genotypic variance for crude protein in this population (Table 3.1). The heritability estimate for protein using four environments with three replications (0.93) or even three environments with four replications (0.92) would be similar to the five environment/two replication combination used in our study (Table 3.4).

For lysine and threonine, the combination of five environments and two replications per environment used in our study was sufficient to detect a 2.5% difference in the overall RIL mean (1.6 g kg^{-1} cp for Lys and 0.9 g kg^{-1} cp for Thr) for these traits (Table 3.2), but not a 1.25% difference. For Lys, the same magnitude of difference could be detected by using only four testing environments and two replications. In addition, detection at the 2.5% level can be attained for Lys and Thr by reducing the number of environments to three and adding two replications (total of four replications/environment). This would be of interest to breeders in a situation where adding additional environments is not cost-effective within the breeding program. Again, the reduction in environments and replications reduces the heritabilities of

these traits greatly. A reduction from 10 plots to two plots results in a reduction in heritability from 0.69 to 0.30 and from 0.86 to 0.55 for Lys and Thr, respectively (Table 3.4). For both of these traits, similar heritability values to those obtained for the five environments/two replications could be attained by growing tests in only four environments and adding an additional replication.

To detect 2.5% of the difference between genotypes for Met, Cys, and Met+Cys, our current allocation of replications and environments would not be suitable, though the addition of two replications in each environment (total of five environments and four replications/environment) would allow this level of detection for all three of these traits. With five environments and four replications/environment, we could detect differences of 0.4, 0.4.and 0.7 g kg^{-1} cp for Met, Cys, and Met+Cys, respectively (Table 3.2, Table 3.3). Differences in these traits at the 2.5% level would go undetected in our current layout, but the use of two replications within seven environments would achieve this level of precision (Table 3.3). The detection of Met or Cys, at the 1.25% level would require up to 16 environments even with four replications per environment. On the other hand, a difference of 1.25% could be detected for Met+cys using eight environments and four replications. Estimates of heritability for these traits were reduced to extremely low levels (0.21 to 0.31) in a single environment with four replications (Table 3.4). Selection in three environments and four replications would result in heritabilities of 0.44, 0.57, and 0.51 for Met, Cys, and Met+Cys, respectively. The proper utilization of resources will vary depending on the specific sulfur-containing amino acid based on these heritability estimates.

Correlation between traits is extremely important in regard to selection by breeders in terms of the effect of selection on correlated traits. It is of interest to determine the effect of

selection for protein and various amino acids on important agronomic traits. Phenotypic correlations, based on means for each trait over all environments for the 140 RILs are presented in Table 3.5. Crude protein (r = -0.50) was negatively associated with seed yield while Lys/cp (r = 0.49), and Thr/cp (r = 0.49) were positively related to yield in our population. Met was positively (r = 0.19) associated with yield, while Cys/cp and Met+Cys/cp were not. None of these traits were correlated at a level which would cause a serious concern in selection. The negative correlation detected in our population between crude protein and yield has been detected previously (Brim, 1973; Kwon and Torrie, 1964; Shorter et al., 1977; Wilcox and Guodong, 1997). This may be associated with the challenge of maintaining protein levels as seed yields have continued to increase over the past several decades in the USA (Wilson, 2004).

The association between protein or amino acids and maturity were negligible (r = -0.24 to 0.22) from a breeding and selection standpoint, though they were significant (P = 0.05). The negative correlation between crude protein and maturity was somewhat surprising as later maturing genotypes have been associated with higher protein in other populations (Simpson and Wilcox, 1983). Correlations between seed weight and crude protein, Lys/cp, Thr/cp, and Met/cp were significant ($P \leq 0.01$), but the relationships with seed yield would only be explained between 13 to 15% of the variation in amino acid concentrations. Crude protein and seed weight were negatively associated in our population (r = -0.36). Previously, seed weight has been associated with increased protein in other populations (Fehr and Weber, 1968; Kwon and Torrie, 1964). No associations were detected between protein or amino acids and plant height or lodging.

Significant G×E interaction was detected in our Benning × Danbaekkong population for crude protein, Lys, Thr, Met, Cys, and Met+Cys when it was grown in five environments. Even

with this G×E interaction, genotypic differences were found for protein and each of the amino acids evaluated in our study. Overall, the impact of G×E detected in this population appears to be minimal based on the relative components of variance for these quality traits, which is in accordance with past studies. The optimum allocation of resources is described for crude protein, Lys, Thr, Met, Cys, and Met+Cys and provides a framework for breeders in terms of the most cost-effective manner by which to select for these traits. Moreover, the expected response to selection for these traits can be determined from our data for various selection units.

References

Allen F., Rasmusson R. (1978) Optimal Environments for Yield Testing1. Crop Science 18:747.

Bernardo R. (2002) Breeding for quantitative traits in plants Stemma Press. Woodbury, MN.

Boerma H., Hussey R., Phillips D., Wood E., Rowan G., Finnerty S. (1997) Registration of'Benning'soybean. Crop science (USA).

Boerma H.R., D.A. Ashley, S.A. Harrison. (1985) Sampling variation and resource allocation in soybean canopy photosynthesis experiments. Crop Science 25:967-970.

Brim C. (1966) A modified pedigree system of selection in soybeans. Crop Sci 6:220.

Brim C. (1973) Quantitative genetics and breeding. Soybeans: improvement, production, and uses. Madison: American Society of Agronomy:155-186.

Brim C., Burton J. (1979) Recurrent selection in soybeans. II. Selection for increased percent protein in seeds. Crop Science 19:494.

Brummer E., Graef G., Orf J., Wilcox J., Shoemaker R. (1997) Mapping QTL for seed protein and oil content in eight soybean populations. Crop Science 37:370-377.

Burton J., Wilson T. (1998) Registration of 'Prolina'soybean. Crop Science 39:294.

Byth D.E., Weber C.R., Caldwell B.E. (1969) Correlated Truncation Selection for Yield in Soybeans1. Crop Sci. 9:699-702. DOI: 10.2135/cropsci1969.0011183X000900060007x.

Carter Jr T.E., J.W. Burton, J.J. Cappy, D.W. Israel, and H.R. Boerma. (1983) Coefficients of variation, error variances, and resource allocation in Soybean growth analysis experiments. Agronomy Journal 75:691-696.

Cherry J., Bishop L., Hasegawa P., Leffler H. (1985) Differences in the fatty acid composition of soybean seed produced in Northern and Southern areas of the USA= Différences de

composition en acides gras des graines de soja produites dans le Nord et dans le Sud des Etats-Unis. Phytochemistry 24:237-241.

Dornbos D., Mullen R. (1992) Soybean seed protein and oil contents and fatty acid composition adjustments by drought and temperature. Journal of the American Oil Chemists' Society 69:228-231.

East E. (1916) Studies on size inheritance in Nicotiana. Genetics 1:164.

Erickson L.R., Beversdorf W.D., Ball S.T. (1982) Genotype × Environment Interactions for Protein in Glycine max × Glycine soja crosses1. Crop Sci. 22:1099-1101. DOI: 10.2135/cropsci1982.0011183X002200060003x.

Fehr W. (1968) Mass Selection by Seed Size and Specific Gravity in Soybean Populations1. Crop Science 8:551.

Fehr W., Caviness C., Burmood D., Pennington J. (1977) Stage of soybean development. Ames: Iowa State University.

Fehr W., Hoeck J., Johnson S., Murphy P., Nott J., Padilla G., Welke G. (2003) Genotype and environment influence on protein components of soybean. Crop Science 43:511-514.

Gibson L., Mullen R. (1996) Soybean seed composition under high day and night growth temperatures. Journal of the American Oil Chemists' Society 73:733-737.

Hanson W. (1963) Heritability.

Hühn M. (1996) Non-parametric analysis of genotype× environment interactions by ranks. Genotype-by-environment interaction:235—271.

Kang M. (1997) Using genotype-by-environment interaction for crop cultivar development. Advances in Agronomy 62:199-252.

Kim S.-D., E.-H. Hong, Y.-H. Kim, S.-H. Lee, Y.-K. Seong, K.-Y. Park, Y.-H. Lee, Y.-H. Hwang, E.-H. Park, H.-S. Kim, Y.-H. Ryu, R.-K. Park, and Y.-S. Kim. (1996) A new high protein and good seed quality soybean variety "Danbaegkong". RDA Journal of Agricultural Science- Upland and Industrial Crops 38:228-232.

Kwon S., Torrie J. (1964) Heritability and interrelationship among traits of two soybean populations. Crop Sci 4:196-198.

Mendenhall W., Scheaffer R.L., Wackerly D.D. (1973) Mathematical statistics with applications Duxbury Press Boston.

Miller J. (1979) Direct and Indirect Recurrent Selection for Protein in Soybeans1. Crop Science 19:101.

Nyquist W., Baker R. (1991) Estimation of heritability and prediction of selection response in plant populations. Critical Reviews in Plant Sciences 10:235-322.

Primomo V., Falk D., Ablett G., Tanner J., Rajcan I. (2002) Genotype× environment interactions, stability, and agronomic performance of soybean with altered fatty acid profiles. Crop Sci 42:37.

Reese P., Hussey H. (1988) Resource allocation in experiments measuring soybean tolerance to soybean cyst nematode. Crop Science 28:589.

SAS. (2003) SAS/STAT User's Guide, Version 9.1.

Sato K., Ikeda T. (1979) The growth responses of soybean plant to photoperiod and temperature, 4: The effect of temperature during the ripening period on the yield and characters of seeds. Japanese Journal of Crop Science (Japan).

Scapim C., Oliveira V., Braccini A., Cruz C., Andrade C., Vidigal M. (2000) Yield stability in maize (Zea mays L.) and correlations among the parameters of the Eberhart and Russell, Lin and Binns and Huehn models. Genetics and Molecular Biology 23:387-393.

Schutz W. (1967) Genotype Environment Interactions in the Regional Testing of Soybean Strains1. Crop Science 7:125.

Shannon J., Wilcox J., Probst A. (1972) Estimated gains from selection for protein and yield in the F4 generation of six soybean populations. Crop Science 12:824.

Shorter R., Byth D., Mungomery V. (1977) Genotype× environment interactions and environmental adaptation. II. Assessment of environmental contributions. Australian Journal of Agricultural Research 28:223-235.

Simpson A., Wilcox J. (1983) Genetic and phenotypic associations of agronomic characteristics in four high protein soybean populations. Crop Science 23:1077.

Smith R.R., Weber C.R. (1968) Mass Selection by Specific Gravity for Protein and Oil in Soybean Populations1. Crop Sci. 8:373-377. DOI: 10.2135/cropsci1968.0011183X000800030033x.

Vollmann J., Fritz C.N., Wagentristl H., Ruckenbauer P. (2000) Environmental and genetic variation of soybean seed protein content under Central European growing conditions. Journal of the Science of Food and Agriculture 80:1300-1306.

Weber C., Fehr W. (1970) Registration of Provar soybeans. Crop Sci 10:728.

Weber W., Wricke G., Westermann T. (1996) Selection of genotypes and prediction of performance by analysing genotype by environment interactions. Genotype Environ. Inter 13:354—370.

Wilcox J. (1998) Increasing seed protein in soybean with eight cycles of recurrent selection. Crop Science 38:1536.

Wilcox J., Cavins J. (1995) Backcrossing high seed protein to a soybean cultivar. Crop Science 35:1036.

Wilcox J., Guodong Z. (1997) Relationships between seed yield and seed protein in determinate and indeterminate soybean populations. Crop Science 37:361.

Wilson R. (2004) Seed composition. Soybeans: improvement, production, and uses 3:621–677.

Wolf R., Cavins J., Kleiman R., Black L. (1982) Effect of temperature on soybean seed constituents: oil, protein, moisture, fatty acids, amino acids and sugars. Journal of the American Oil Chemists' Society 59:230-232.

Yue G., Roozeboom K., Schapaugh Jr W., Liang G. (1997) Evaluation of soybean cultivars using parametric and nonparametric stability estimates. Plant Breeding 116:271-275.

Table 3.1. Combined analysis of variance and estimates of variance components for crude protein and various amino acids for 140 RILs evaluated in five environments.

		Crude Protein				Lysine (Lys)			
Source	df	Mean square	Variance component	Percent of total	F-value	Mean square	Variance component	Percent of total	F-value
	no.	(g kg^{-1} cp)	(g kg^{-1}cp)	(%)		(g kg^{-1}cp)	(g kg^{-1}cp)	(%)	
Env	4	32187	100	16	<.0001	483.4	2	48	<.0001
Rep(env)	5	3300	20	3		13.1	.08	2	
Sets	2	14536	20	3	<.0001	14.4	0	0	<.0001
Env × Sets	8	2286	16	3	<.0001	17.7	0.1	2	<.0001
Rep x Sets(Env)	10	741	13	2		3.3	.05	1	
Geno(Sets)	137	2959	280	44	<.0001	5.7	0.4	10	<.0001
Env × Geno(Sets)	548	220	39	6	<.0001	1.8	0.4	10	<.0001
Pooled error	669	145	145	23		1.0	1.06	26	
		Threonine (Thr)				Methionine (Met)			
Env	4	396.7	1.0	33	<.0001	9.2	0.03	10	<.0001
Rep(env)	5	8.4	.05	2		0.3	0.001	0.3	
Sets	2	21.5	0.2	7	<.0001	0.7	0	0	0.0226
Env × Sets	8	7.6	0.5	17	<.0001	2.0	0.02	7	<.0001
Rep x Sets(Env)	10	2.6	.04	1		0.14	0	0	
Geno(Sets)	137	5.3	0.5	17	<.0001	0.40	0.02	7	<.0001
Env × Geno(Sets)	548	0.78	0.2	7	<.0001	0.23	0.03	10	0.0007
Pooled error	669	0.5	0.5	17		0.18	0.18	66	
		Cysteine (Cys)				Met+Cys			
Env	4	41.4	0.14	34	<.0001	53.8	0.17	18	<.0001
Rep(env)	5	0.4	0	0		0.5	0	0	
Sets	2	10.5	0.02	5	<.0001	16.2	0.02	2	<.0001
Env × Sets	8	3.2	0.03	7	<.0001	6.2	0.06	6	<.0001
Rep x Sets(Env)	10	0.79	0.01	2		1.2	0.009	0.1	
Geno(Sets)	137	0.50	0.03	7	<.0001	1.5	0.08	8	<.0001
Env × Geno(Sets)	548	0.22	0.03	7	<.0001	0.7	0.11	12	<.0001
Pooled error	669	0.16	0.15	37		0.5	0.5	53	

Table 3.2. Mean values for protein and various amino acids, coefficients of variation, and values of traits for a 1.25, 2.5, 5.0, and 10.0% difference between the RIL means.

Trait	Mean	Coefficient of variation	Magnitude of genotype difference expressed as percent of RIL $\bar{X}$			
			1.25%	2.5%	5%	10%
	($g\ kg^{-1}$)	%	$g\ kg^{-1}$			
Crude protein	468.3	2.6	5.9	11.7	23.4	46.8
Lys/cp	63.3	1.6	0.8	1.6	3.2	6.3
Thr/cp	36.1	1.9	0.5	0.9	1.8	3.6
Met/cp	14.2	3.0	0.2	0.4	0.7	1.4
Cys/cp	15.3	2.6	0.2	0.4	0.8	1.5
Met+Cys/cp	29.6	2.4	0.4	0.7	1.5	2.9

Table 3.3. Number of replications and environments to detect 1.25, 2.5, 5.0 and 10% of the difference between two genotype means.

Trait	Magnitude of difference	Replications 2	3	4
Protein	1.25% of $\bar{X}$	29	23	10
	2.5% of $\bar{X}$	8	6	3
	5.0% of $\bar{X}$	2	2	1
	10% of $\bar{X}$	1	1	1
Lysine	1.25% of $\bar{X}$	13	11	10
	2.5% of $\bar{X}$	4	3	3
	5.0% of $\bar{X}$	1	1	1
	10% of $\bar{X}$	1	1	1
Threonine	1.25% of $\bar{X}$	14	12	10
	2.5% of $\bar{X}$	5	4	3
	5.0% of $\bar{X}$	1	1	1
	10% of $\bar{X}$	1	1	1
Methionine	1.25% of $\bar{X}$	28	21	17
	2.5% of $\bar{X}$	7	6	5
	5.0% of $\bar{X}$	3	2	2
	10% of $\bar{X}$	1	1	1
Cysteine	1.25% of $\bar{X}$	25	19	16
	2.5% of $\bar{X}$	7	5	4
	5.0% of $\bar{X}$	2	2	1
	10% of $\bar{X}$	1	1	1
Met+Cys	1.25% of $\bar{X}$	21	16	8
	2.5% of $\bar{X}$	7	6	3
	5.0% of $\bar{X}$	2	2	1
	10% of $\bar{X}$	1	1	1

Table 3.4. Variance component heritability estimates for various combinations of environments and replications for crude protein and various amino acids.

Env/Rep	Crude Protein	Lys/cp	Thr/cp	Met/cp	Cys/cp	Met+cys/cp
Two replications						
1 environment (2 plots)	0.80	0.30	0.55	0.14	0.22	0.18
2 environments (4 plots)	0.83	0.46	0.70	0.31	0.36	0.30
3 environments (6 plots)	0.88	0.56	0.78	0.33	0.46	0.40
4 environments (8 plots)	0.90	0.63	0.82	0.40	0.54	0.47
5 environments (10 plots)	0.93	0.69	0.86	0.45	0.59	0.53
Three replications						
1 environment (3 plots)	0.75	0.35	0.64	0.18	0.27	0.22
2 environments (6 plots)	0.86	0.52	0.75	0.31	0.43	0.36
3 environments (9 plots)	0.90	0.62	0.82	0.40	0.54	0.47
4 environments (12 plots)	0.93	0.68	0.85	0.47	0.60	0.53
5 environments (15 plots)	0.94	0.74	0.89	0.52	0.65	0.59
Four replications						
1 environment (4 plots)	0.78	0.38	0.63	0.21	0.31	0.25
2 environments (8 plots)	0.88	0.56	0.78	0.35	0.47	0.40
3 environments (12 plots)	0.92	0.65	0.84	0.44	0.57	0.51
4 environments (16 plots)	0.94	0.71	0.88	0.53	0.64	0.57
5 environments (20 plots)	0.95	0.76	0.90	0.57	0.70	0.61

Table 3.5. Phenotypic correlation coefficients between amino acid concentrations and agronomic traits of 140 RILs in the Benning × Danbaekkong population across five environments.

Trait[†]	Crude Protein	Lys/cp	Thr/cp	Met/cp	Cys/cp	Met+Cys/cp
Seed yield	-0.50**	0.49**	0.49**	0.19*	-0.01	0.08
Maturity	-0.24*	0.21*	0.22*	0.07	-0.22*	-0.10
Seed weight	-0.36**	0.39**	0.35**	0.37**	0.13	0.25
Plant height	-0.11	0.07	0.09	-0.01	-0.07	0.01
Lodging	-0.13	0.06	0.08	0.13	-0.07	0.01

*, ** significant at the $P \leq 0.05$ and $P \leq 0.01$ level of significance

[†]Seed yield is based on four environments and two replications/environment, maturity on three environments and two replications/environment, seed weight on on two environments and two replications/environment, and plant height and lodging on three environments and two replications/environment.

CHAPTER 5

CONCLUSIONS

Due to the importance of soybean as a provider of high-quality protein in livestock production, soybean breeders are interested in gaining a better understanding of the genetic basis and environmental influences on the amino acids which compose this protein. In this study, we conducted one of the first QTL mapping studies for several important amino acids in soybean and also quantified the optimum allocation of resources in breeding for crude protein, Lys/cp, Thr/cp, Met/cp, Cys/cp, and Met+Cys/cp based on significant genotype × environment (G×E) interaction detected across five field environments.

Near-infrared reflectance spectroscopy (NIR) was used to phenotype 140 recombinant inbred lines for crude protein and amino acid content. A number of QTL were detected in our study for crude protein and selected amino acids. A major protein QTL on chr 20 (Lg-I) was found to explain a large proportion of the phenotypic variance for crude protein, Lys/cp, Thr/cp, Met/cp, and Met+Cys/cp based on composite interval mapping (CIM). The allele from the 'Danbaekkong' parent at this QTL resulted in increased crude protein content but reduced levels of each amino acid. CIM also identified several minor QTL for each amino acid on various chromosomes. By gaining an understanding of the location and effects of QTL, breeders can introgress specific alleles in combination to improve traits of interest by selecting genotypes based on DNA markers tightly linked to QTL. It will be of interest in future research to determine if the QTL detected on chr 20 is composed of a single gene controlling each of these traits (pleiotropic effect) or a number of tightly linked genes uniquely controlling concentrations

of each amino acid. This QTL has previously been fine-mapped and efforts to clone this gene are underway. The improvements in phenotyping and genotyping precision and efficiency can also be observed from our study. The use of near-infrared reflectance spectroscopy (NIR) for phenotyping and the Illumina GoldenGate Assay for genotyping the RILs with 1,536 SNP markers in the soybean genome shows that technogical advances are rapidly advancing and aiding breeding efforts greatly.

The second goal of this research was to determine the magnitude of G×E interaction for crude protein, Lys/cp, Thr/cp, Met/cp, Cys/cp, and Met+Cys/cp and its effect on how resources should be allocated in a program breeding for these traits. A population of RILs from the Benning × Danbaekkong cross was grown in five environments and traits were assessed. It appears that G×E interactions in this population would be of minor concern to breeders aiming to improve these traits based on the relative magnitudes of the genotypic and G×E variance components. The optimum level of discrimination among genotypes for these traits was also determined. The data show that more plots (environments/replications) are necessary to discriminate between small differences in genotypic means (1.25%) for crude protein, Met/cp, Cys/cp, and Met/cp relative to Lys/cp and Thr/cp. The required level of precision would be left to the discretion of individual breeders. The nature and magnitude of G×E interaction interferes with the breeders' ability to select for traits of interest when the goal is to breed for broad trait adaptation. The goal of this study was to provide the breeding community an idea of how these traits are affected by the environment and how breeding efforts may proceed in the presence of G×E interactions.

Our results suggest that a high-protein cultivar with enhanced amino acid characteristics could be developed. At this time, it may be of greater interest to develop a cultivar with

moderate protein levels and increased levels of critical amino acids, and to avoid sacrificing yield. Marker-assisted selection (MAS) could aid this endeavor greatly. It is not surprising to see so little variability for the sulfur-containing amino acids Met and Cys between Danbaekkong and the elite checks. This has been noted in the past as the limitation in terms of increasing these two amino acids in soybean cultivars. Our findings concerning G×E interaction for these traits are promising in that all of the time and effort necessary to create improved cultivars for these traits should not be greatly undermined by environmental effects.