This is to certify that the

thesis entitled

EVALUATION OF QTL ALLELES FROM THE WILD GLYCINE SOJA THAT INCREASE PROTEIN CONTENT IN GLYCINE MAX

presented by

Audrey M. Sebolt

has been accepted towards fulfillment
of the requirements for

MS degree in CSS

James D. Kelly
Major professor

Date 5-10-99

EVALUATION OF QTL ALLELES FROM WILD *GLYCINE SOJA* THAT INCREASE PROTEIN CONTENT IN *GLYCINE MAX*

By

Audrey M. Sebolt

A THESIS

Submitted to
Michigan State University
in partial fulfillment of the requirements
for the degree of

MASTERS

Crop and Soil Sciences

1999

ABSTRACT

EVALUATION OF QTL ALLELES FROM WILD *GLYCINE SOJA* THAT INCREASE PROTEIN CONTENT IN *GLYCINE MAX*

By

Audrey M. Sebolt

Genes from *Glycine soja* that increase protein content were first evaluated to determine if they would stably increase protein content in a backcross (BC) population. Provided a consistent response was observed, the *G. soja* genes would be analyzed in the genetic backgrounds of soybean cultivars 'Parker', 'Kenwood', and C1914. RFLP analyses were conducted using markers that map to regions on Linkage Groups (LG) E and I, where the high protein genes from *G. soja* mapped. The *G. soja* allele on LG I increased protein content after it was backcrossed into a *G. max* background. Lines homozygous for this *G. soja* allele exhibited a 174 kg ha^{-1} yield reduction when compared to the *G. max* allele.

One line, selected from this BC population, was crossed to Kenwood, Parker, and C1914. The *G. soja* allele on LG I increased protein content in the backgrounds of Kenwood and Parker but not in the C1914 background. These data suggest that the high protein gene in *G. soja* gene is allelic to an allele in C1914 but not in Parker or Kenwood. The high protein allele from *G. soja* was associated with less yield in the Kenwood and Parker crosses but not in the C1914 cross.

ACKNOWLEDGMENTS

I would like to first thank my committee members Drs. Brian Diers, James Kelly, and Amy Iezzoni for their support and guidance. A special thanks to Dr. James Kelly for agreeing to be my advisor at the eleventh hour. I would also like to extend my gratitude to John Boyse for assisting me with my fieldwork. Thank you Deane Lehmann for your moral and technical support, and friendship. Not only did she teach me the finer points of proper lab technique (try not to break the glassware and a mop is very handy) but also helped me through the stressful times.

Special thanks to my "stat" buddies Kirsten Jaglo-Ottosen (the other blonde), Maria Theresa Ospina, and Suzanne Downey. Not only did they help me through two semesters of stats, but also through this entire experience of finishing my thesis. Thanks also to Ann Gustafson, Holly Little, and Sarah Gilmour, my running buddies, for providing an outlet for accumulated stress these last few months.

I thank my father and mother for instilling in me an appreciation for agriculture. But most of all, many thanks to my husband, Chris, who has been there since the beginning and traveled this road with me. Thank you for all of your love and support when I needed it most!

TABLE OF CONTENTS

LIST OF TABLES

Chapter 2:

Chapter 3:

LIST OF FIGURES

Chapter 2

CHAPTER 1

LITERATURE REVIEW

Introduction

Prior to the 1920's, soybean [*Glycine max*, (L.) Merr.] was imported into the U.S. due to the lack of interest in the production of this field crop. In the mid-1920's, the discovery that soybean meal provided a high protein supplement for livestock and poultry feed led to an increase in the demand for soybean. Thereafter, soybean acreage in the U.S. rapidly increased (Smith and Huyser, 1987).

Soybean breeders have attempted to achieve higher seed protein of elite cultivars with limited success. The lack of diversity in elite populations has hindered this effort and genes for increased protein content have been sought elsewhere. Wild populations are one source of "new" genetic diversity and genetic mapping has been useful in identifying genes for increased protein content in the progeny of interspecific crosses.

In the following pages, researchers attempts to achieve higher seed protein are reviewed. Whether genes that increase seed protein were successfully transferred from *G. soja* to existing soybean cultivars will be examined as well. In

addition to these topics, the use of molecular techniques and the progress of genetic mapping in soybean will also be discussed.

Conventional Plant Breeding

Approximately 4.6% of the amino acid composition soybean protein consists of methionine and cysteine (Wilson, 1987). Unfortunately, the amount of these amino acids in soybean does not meet the nutritional recommendation set by FAO and thus, some foods that contain soy-protein may need to be supplemented (Wilson, 1987). Because these amino acids are important for animal nutrition, increasing the level of methionine and cysteine in soybean will increase its nutritional value. At this time, no breeding efforts have successfully altered the level of these amino acids, in part, due to the expense of the assays and the time involved to evaluate lines for amino acid concentration. Breeding programs simply attempt to increase seed protein by using traditional plant breeding methods as a means to increase total seed protein concentration. While this attempt has been made by numerous breeders during the last 20 years, conventional plant breeding has unfortunately not increased percent methionine and cysteine (Burton et al, 1982) of the soybean seed.

Researchers who first initiated the process of increasing seed protein in soybean, used experimental breeding lines as a source of genes for increased protein content. In a comparison of direct and indirect selection, Miller and Fehr (1979) developed two populations with the goal of increasing protein

concentration. Whereas indirect selection was used to increase protein content by selecting lines with low oil, direct selection was used as a means to select lines with high protein.

Direct selection resulted in an increase in seed protein two times greater than detected through indirect selection. Despite the improvement in seed protein, there was a significant decrease in oil concentration when direct selection was attempted. Because both protein meal and vegetable oil from soybean are of value, to increase both or to increase one without decreasing the other would be advantageous (Miller and Fehr, 1979). Indirect selection was shown to be more effective in increasing seed protein content without reducing yield.

Brim and Burton (1979) used recurrent selection as a means to increase seed protein content. After five cycles of recurrent selection, seed protein content was increased by 10.2% higher than the base population in one of the four populations studied. Despite this success, three of the four populations exhibited decreased yields after cycle five of recurrent selection. In addition, negative correlations between protein and oil concentration and seed protein and yield were observed.

Plant breeders have also used the backcross (BC) method to increase seed protein of soybean. The BC method is used to transfer a desirable gene or group of genes from one plant into another that has an overall good agronomic background (Allard, 1960). After the initial cross, successive backcrosses to the recurrent parent are made to recover the characteristics of the recurrent parent

while maintaining the gene or group of genes from the donor parent. Lines must then be selfed after the final BC in order for the gene (or genes) that were transferred to become homozygous (Allard, 1960). The total number of backcrosses is dependent upon the crop and the objective of the breeding program and the heritability of the trait.

The backcross breeding method has been used for over a century. When working with small grain crops, Harlan and Pope (1922) expressed a preference for the BC breeding method because the desired character transferred was fixed; in other words, the character was not apt to easily segregate out of the population. In addition, the BC method is considered of value because morphological and agronomical features of the improved variety can be predicted in advance and there is a high degree of genetic control of a population (Allard, 1960).

Utilizing the BC method, Wehrmann et al. (1987) attempted to transfer genes that increase protein content from the *G. max* accession Pando to three high yielding soybean cultivars. Seed yields, equivalent to the recurrent parent, were recovered after only two backcrosses; however, there was limited success in increasing seed protein content. Pando was recorded as containing 480 g kg^{-1} seed protein. The three high yielding recurrent parents had 361, 362, and 413 g kg^{-1} seed protein content while the three highest yielding BC_2 F_2-derived lines had only slightly greater seed protein content averaging 373, 376, and 427 g kg^{-1}. The insignificant increase may have been attributed to difficulties in identifying a

donor parent that could successfully transfer loci controlling high seed protein to progenies of consecutive backcrosses (Wilcox and Cavins, 1995).

With the use of the BC method, Wilcox and Cavins (1995) also attempted to transfer genes for increased seed protein content from Pando to the high yielding cultivar Cutler 71. With each BC, rapid progress was made in recovering yield and other agronomic traits, while maintaining high seed protein. Results, which were similar to Wehrmann et al. (1987), also suggested that not all of the genes controlling seed protein contents in Pando were transferred to the recurrent parent. The protein content of Pando was recorded as 498 g kg^{-1} however, the line with the highest seed protein content was only 472 g kg^{-1}. Despite the significant increase in protein content, there was no dramatic decrease in yield; BC_3 progenies were equal to or greater in seed yield than Cutler 71.

Pleiotropic effects may explain the inverse relationship between seed protein and oil contents found in the previous studies (Graef et al., 1989). However, the basis for decreased yields when seed protein contents are significantly increased may be more complex. If genes controlling the trait studied are linked to yield genes, there could be difficulties in separating the two traits, and to further complicate matters, yield is generally a polygenic trait. In the populations they examined, Brim and Burton (1979) attributed the negative correlation between seed yield and protein content to pleiotropy, tight linkages, or both.

Hartwig and Hinson (1972) were successful in transferring genes from the high protein parent D60-7965 to Bragg, a high yielding cultivar, with out a reduction of yield. After two backcrosses, one line was equivalent in seed yield to Bragg and seed protein content of D60-7965. They concluded "that high protein genes *per se* did not significantly influence seed yield." Despite a significant negative correlation between yield and seed protein content in the BC_1 lines, there was a positive correlation in BC_2 lines. This was because lines were reduced from 25% donor germplasm in the BC_1 generation to 12.5% in the BC_2 generation, thus decreasing the number of deleterious genes that may have contributed to the decrease in yield.

Glycine soja

The wild progenitor of *G. max* is *G. soja* and both are diploid annual species (2n=40x), however, their plant morphology differs dramatically (Hymowitz and Singh, 1987). *G. max*, which has never been found in the wild, exhibits an upright, sparsely branched, bush-type growth habit, with seeds weighing 10 to 20 g $(100 \text{ seeds})^{-1}$. In contrast, *G. soja* is known for its undesirable characteristics such as its susceptibility to lodging, vining, and colored seed coats (Weber, 1950; Carpenter and Fehr, 1986). The species has a tendency to shatter, grows prostrate with the ground, and seeds weigh roughly 0.1 g $(100 \text{ seeds})^{-1}$. *G. soja* is found in the wild and is distributed throughout China and adjacent areas

such as Korea, Japan, Taiwan, and some countries of the former Soviet Union. In the past, *G. soja* was referred to as *G. ussuriensis* (Hymowitz and Singh, 1987).

Sources of genes that increase protein content in elite U.S. soybean cultivars are limited due to the lack of diversity present in this germplasm. U.S. soybean cultivars derived from the hybridization of plant introductions (PI; Keim et al., 1989) and few accessions have made large genetic contributions to the pedigrees of elite cultivars. It is estimated that only ten accessions contribute 88% to the Northern U.S. germplasm (Delanney et al., 1983; Fehr, 1987). Furthermore, *G. max* contains less genetic diversity than its progenitor *G. soja* because of bottlenecks during the domestication process. These bottlenecks resulted in a loss of alleles during domestication and further losses occurred through modern plant-breeding practices (Tanksley and McCouch, 1997). The reduced genetic variation in elite germplasm, because of these bottlenecks, has resulted in a slow rate of genetic improvement by plant breeders. When breeding populations have low genetic variation, breeders are not likely to identify new and useful gene combinations (Tanksley and Nelson, 1996). In order to recapture lost alleles, a plant breeder should consider the wild ancestors of crop species as a source of "new" genes.

Weber (1950) published a study in which a segregating population was created by a cross between *G. soja* and *G. max*. Results showed there was a strong negative correlation between percent protein and percent oil in the population, which is consistent with findings in *G. max*. Weber (1950) also

indicated that if a breeder attempted to transfer high protein content from the *G. soja* parent to *G. max*, there would be difficulty in recovering seed size and oil content. Harlan (1976) reported that transferring genes that increase seed protein content from *G. soja* into *G. max* should be possible and the deleterious characteristics of *G. soja* can be selected against in a backcrossing program. Ultimately, yields could be recovered that are similar to the recurrent parent.

Using *G. soja* in breeding programs is simplified because *G. soja* and *G. max* are interfertile, therefore, genes from *G. soja* can easily be transferred to *G. max* through crossing. To obtain lines similar to the recurrent *G. max*, Ertl and Fehr (1985) found that three backcrosses to *G. max* were required. Carpenter and Fehr (1986) confirmed that three backcrosses were needed and resistance to lodging and absence of vining were most difficult to recover in early BC generations.

Molecular Analyses

Molecular tools such as isozyme or restriction fragment length polymorphism (RFLP) markers assist in the mapping of genes and the selection of lines that contain transferred genes. Genetic markers can indicate genetic diversity through the selection of marker alleles from the wild species. Molecular markers have also proven beneficial in helping to decrease the number of lines evaluated in the field because continuous selection of breeding lines can be imposed as lines are advanced (Suarez et al., 1991). Despite these benefits, this process is often

more costly when compared to traditional plant breeding (Tanksley and Nelson, 1996) because different resources and expertise are required compared to traditional plant breeding.

Genes that control quantitative traits are referred to as quantitative trait loci (QTL) (Hartl, 1994). Manipulation of QTLs could eventually lead to superior varieties, however, genes controlling the trait studied must first be mapped with genetic markers (Hartl, 1994). The number of QTLs linked to markers and the amount of recombination between the marker loci and the QTL are critical when selecting for a trait such as increased protein content.

Successful application of marker-assisted selection has been accomplished in tomato. Quantitative trait loci controlling soluble solids (SS) content were mapped in a BC population using genetic markers. The genes that increased SS were transferred from the wild to the cultivated tomato (Osborn et al., 1987). Researchers first screened a high SS derived BC line for variation in RFLPs compared to the low SS recurrent and high SS donor parent. Two cDNA clones that hybridized to RFLPs were identified and the authors concluded that the RFLPs had been introduced into the BC line from the high SS donor parent (Osborn et al., 1987).

To determine if RFLPs identified by the two cDNA clones were linked to a gene(s) that controlled SS content in tomato, Osborn et al. (1987) developed a population in which the previously mentioned high SS derived BC line was crossed to a low SS tomato processing line. Analysis of variance revealed that a

RFLP locus was linked to one or a group of loci affecting SS content. The RFLP allele from the high SS BC derived line was associated with significantly higher SS content suggesting this linkage relationship could be used in a tomato breeding program. Lines with increased SS content could be identified by selecting for the RFLP allele.

Tanksley and Nelson (1996) proposed a new breeding method known as the advanced backcross QTL analysis. Their purpose of the method was to compare QTL analyses of traditional balanced populations with advanced backcross populations. Results, through computer simulations, indicated that genetic mapping should be done no later than the BC_2 or BC_3 generation. Researchers also concluded that the genotype and phenotype of lines in the BC_2 and BC_3 generation resemble the recurrent parent more so than earlier BC generations because the frequency of deleterious or undesirable alleles from the donor parent is diminished.

In soybean, Graef et al. (1989) and Suarez et al. (1991) studied the association between isozyme markers and quantitative traits. Both groups of investigators used the same two *G. max* by *G. soja* populations and found that vegetative traits, including seed protein and seed oil content, were significantly associated with isozyme loci, although, associations were population specific. Suarez et al. (1991) found there were a limited number of polymorphic isozyme loci between the parents of both crosses, a detriment to using marker-assisted

selection. Only six isozyme markers were polymorphic for the first cross and eight for the second cross.

To be useful across environments, QTLs must be found that are stable in different environments. Environmentally sensitive QTLs are present only when environmental conditions are similar to the environment in which the QTL was first identified (Mian et al. 1996). Therefore, environmentally sensitive QTLs lack consistency across environments. Over three locations, Mian et al. (1996) discovered several molecular markers associated with seed weight in two soybean populations. Effective marker-assisted selection for seed weight was feasible and could be applied to other breeding programs because useful QTLs were identified in several populations and environments.

QTL mapping of several traits in interspecific soybean populations has been successful. Keim et al. (1990) mapped QTL for seed hardness using RFLP markers using a population created by a cross between the *G. max* experimental line A81-356022 and the wild *G. soja* accession PI 486916,. Five QTL were identified that combined explained 70% of the variation for seed hardness in the population. Markers mapped to these regions could be used to develop genotypes with varying levels of seed-coat hardness.

The locations of genes that increase protein content in soybean were mapped by Diers et al. (1992) in a population derived from a cross between the *G. max* experimental line A81-356022 and the *G. soja* accession PI 468916. Several markers associated with significant variation for protein content were identified.

The RFLP maker K011 explained 42% of the variation for protein content and this marker mapped to LG I. The most significant markers ($P < 0.01$) mapped to LGs E and I (Shoemaker and Specht, 1995), suggesting that important genes for protein and oil content are located within these linkage groups. Furthermore, all *G. soja* alleles at loci significant for protein content were associated with greater protein content than *G. max* alleles.

Brummer et al. (1997) mapped genes controlling protein and oil content in eight different soybean populations. One particular population exhibited a very strong QTL for protein on LG I. A noteworthy difference between this population and the other seven populations was that one parent for this population was 25% *G. soja*. The RFLP marker A144 was used to detect this strong QTL and the marker explained 27.5% of the genetic variation for protein content in the population. QTL for protein and oil content were also found in other populations, but these QTL were population specific. Data suggested that marker-assisted selection could be used as a tool to pyramid different protein QTLs into a common background, creating a population higher in seed protein content and superior to populations already derived.

Most researchers value the use of interspecific crosses to gain knowledge of where genes are located. Mansur et al. (1993), however, concluded that segregating populations resulting from interspecific crosses should not be used to study agronomic characteristics. Furthermore, populations that result from an interspecific cross can not be evaluated in a meaningful way because these

populations have agronomic characteristics that are not desirable, such as trailing vines and pods that have a tendency to shatter. A breeding population that results from similar phenotypic parents is more ideal for QTL mapping and evaluation of traits such as yield.

Further investigations of populations with *G. soja* and *G. max* as parents should eventually lead to a better understanding of where genes that increase protein content are located and their relationship with other traits. More importantly, a backcross population in which genes from *G. soja* that increase protein content were transferred to *G. max* needs to be conducted to evaluate the stability of *G. soja* genes. Finally, the correlation between seed yield and protein content is an economic issue that, with continued research, could be better understood and possibly resolved.

REFERENCES

Allard, R.W. 1960. Principles of plant breeding. John Wiley and Sons, Inc. New York, N. Y. p. 150-165.

Brim, C.A., and J.W. Burton. 1979. Recurrent selection in soybean. II. Selection for increased percent protein in seeds. Crop Sci. 19:494-498.

Brummer, E.C., G.L. Graef, J. Orf, J.R. Wilcox, and R.C. Shoemaker. 1997. Mapping QTL for seed protein and oil content in eight soybean populations. Crop Sci. 37:370-378.

Burton, J.W., A.E. Purcell, and W.M. Walter, Jr. 1982. Methionine concentration in soybean protein from populations selected for increased percent protein. Crop Sci. 23:744-747.

Carpenter, J.A., and W.R. Fehr. 1986. Genetic variability for desirable agronomic traits in populations containing *Glycine soja* germplasm. Crop Sci. 26:681-686.

Delanney, X., D.M. Rodgers, and R.G. Palmer. 1983. Relative genetic contribution among ancestral lines to North American soybean cultivars. Crop Sci. 23:944-949.

Diers, B.W., P. Keim, W.R. Fehr, and R.C. Shoemaker. 1992. RFLP analysis of soybean seed protein and oil content. Theor. Appl. Genet. 83:608-612.

Ertl, D.S., and W.R. Fehr. 1985. Agronomic performance of soybean genotypes from *Glycine max* X *Glycine soja* crosses. Crop Sci. 25:589-592.

Fehr, W.R. 1987. Breeding methods for cultivar development. p. 249-293. *In* J.R. Wilcox (ed) Soybean: improvement, productions, and uses. 2nd edn. Agron. Monogr. 16. ASA, CSSA, and SSSA, Madison, WI.

Graef, G.L., W.R. Fehr, and S.R. Cianzio. 1989. Relation of isozyme genotypes to quantitative characters in soybean. Crop Sci. 29:683-688.

Harlan, H.V., and M.N. Pope. 1922. The use and value of back-crosses in small-grain breeding. Jour. Heredity 13:319-322.

Harlan, J.R. 1976. Genetic resources in wild relatives of crops. Crop Sci. 16:329-332.

Hartl, D.L. 1994. Genetics. Jones and Bartlett Publishers. Boston, Ma. p. 245-247.

Hartwig, E.E., and K. Hinson. 1972. Association between chemical composition of seed and seed yield of soybean. Crop Sci. 12:829-830.

Hymowitz, T., and R.J. Singh. 1987. Taxonomy and speciation. p. 23-48. *In* J.R. Wilcox (ed) Soybean: improvement, productions, and uses. 2nd edn. Agron. Monogr. 16. ASA, CSSA, and SSSA, Madison, WI.

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

Keim, P., R.C. Shoemaker, and R.G. Palmer. 1989. Restriction fragment length polymorphism diversity in soybean. Theor. Appl. Genet. 77:786-792.

Mansur, L.M., K.G. Lark, H. Kross, and A. Oliveira. 1993. Interval mapping of quantitative trait loci for reproductive, morphological, and seed traits of soybean (*Glycine max* L.). Theor. Appl. Genet. 86:907-913.

Mian, M.A.R., M.A. Bailey, J.P. Tamulonis, E.R. Shipe, T.E. Carter Jr., W.A. Parrott, D.A. Ashley, R.S. Hussey, and H.R. Boerma. 1996. Molecular markers associated with seed weight in two soybean populations. Theor. Appl. Genet. 93:1101-1016.

Miller, J.E., and W.R. Fehr. 1979. Direct and indirect selection for protein in soybean. Crop Sci. 19:101-106.

Osborn, T.C., D.C. Alexander, and J.F. Fobes. 1987. Identification of restriction fragment length polymorphisms linked to genes controlling soluble solids content in tomato fruit. Theor. Appl. Genet. 73:350-356.

Shoemaker, R.C., and J.E. Specht. 1995. Integration of the soybean molecular and classical genetic linkage groups. Crop Sci. 35:436-446.

Smith, K.J., and W. Huyser. 1987. World distribution and significance of Soybean. p. 1-22. *In* J.R. Wilcox (ed) Soybean: improvement, production, and uses. 2nd ed. Agron. Monogr. 16. ASA, CSSA, and SSSA, Madison, WI.

Suarez, J.C., G.L. Graef, W.R. Fehr, and S.R. Cianzio. 1991. Association of isozyme genotypes with agronomic and seed composition traits in soybean. Euphytica 52:137-146.

Tanksley, S.D., and S.R. McCouch. 1997. Seed banks and molecular maps: unlocking genetic potential from the wild. Science 277:1063-1066.

Tanksley, S.D., and J.C. Nelson. 1996. Advanced backcross QTL analysis: a method for the simultaneous discovery and transfer of valuable QTLs from unadapted germplasm into elite breeding lines. Theor. Appl. Genet. 92:191-203.

Weber, C.R. 1950. Inheritance and interaction of some agronomic and chemical characters in an interspecific cross in soybean, *Glycine max* x *G. ussuriencsis*. Ames, Iowa: Research Bulletin 374.

Wehrmann, V.K., W.R. Fehr, S.R. Cianzio, and J.F. Cavins. 1987. Transfer of high seed protein to high-yielding soybean cultivars. Crop Sci. 27:927-931.

Wilcox, J.R., and J.F. Cavins. 1995. Backcrossing high seed protein to a soybean cultivar. Crop Sci. 35:1036-1041.

Wilson, R.F. 1987. Seed metabolism. p. 643-686. *In* J.R. Wilcox (ed) Soybean: improvement, productions, and uses. 2nd edn. Agron. Monogr. 16. ASA, CSSA, SSA, Madison, WI.

CHAPTER 2

EVALUATION OF QTL ALLELES FROM THE WILD *GLYCINE SOJA* THAT INCREASE PROTEIN CONTENT IN *GLYCINE MAX*

Introduction

Soybean [*Glycine max* (L.) Merr.] meal protein and oil are currently the most widely produced, traded, and utilized protein meal and vegetable oil source in the world. The protein meal constitutes approximately sixty percent of the total soybean based products and the market for soybean is more dependent upon meal than oil products (Smith and Huyser, 1987). With the global interest in soybean products, research to improve the protein content of soybean is justified.

Previous research established that genes from the wild *G. soja* could increase protein content in *G. max*. In a population developed by crossing a *G. max* experimental line and a *G. soja* plant introduction (PI), Diers et al. (1992) identified two major quantitative trait loci (QTL) from *G. soja* that increase protein content. The two QTL were mapped with restriction fragment length polymorphism (RFLP) markers on Linkage Groups (LG) E and I of the soybean linkage map (Shoemaker and Specht, 1995). Brummer et al. (1997) also mapped a QTL for increased protein content to the same region of LG I as Diers et al. (1992). The population used by Brummer et al. (1997) had one parent that was

25% *G. soja*, suggesting that their high protein gene also came from *G. soja*.

Indeed, there has been several attempts to increase soybean protein content in existing populations, unfortunately researchers have found a negative association between seed protein concentration, yield, and oil concentration (Hartwig and Hinson, 1972; Miller and Fehr, 1979; Brim and Burton, 1979). A better understanding of the interaction of these traits may improve breeding strategies. Breeders may eventually increase protein or oil content while simultaneously increasing yield, using marker-assisted selection (Brummer et al., 1997).

The objective of this study was to analyze regions from *G. soja* on LG E and I to determine if they will stably increase protein content in a backcross (BC) population. The relationship of agronomic traits were also examined with markers that map to these regions.

Materials and Methods

The genetic population used in this study was obtained from the population derived from the interspecific cross between the *G. soja* accession PI 468916 and the *G. max* experimental line A81-356022 developed by Diers et al. (1992). For our study, one line was selected from this original F_2 population because it was homozygous for the *G. soja* regions associated with increased protein content on LGs E and I. This line was the donor parent for three successive backcrosses to A81-356022. During the backcrosses, the *pb* gene, a gene conferring pubescence

tip (Palmer and Kilen, 1987), on LG E and the RFLP marker A144 on LG I were used as selection criteria to recover the *G. soja* regions. The $BC_3 F_1$ plant, that had both regions based on the RFLP marker and *pb*, was selfed and the population was inbred to the F_4 generation through single seed descent to develop F_4 derived lines.

For each BC_3, F_4-derived line, leaf tissue was collected from several plants and DNA extractions were conducted according to Kisha et al. (1997) and Southern blotting, hybridization, and autoradiography were performed as described by Diers and Osborn (1994). A total of ten soybean RFLP markers (Figure 1) were screened against parental DNA and digested with five restriction endonucleases (*Eco*RI, *Hind*III, *Eco*RV, *Dra*I and *Taq*I) to identify polymorphisms.

Fifty-three BC_3, F_4-derived lines were evaluated during the summers of 1996 and 1997 near East Lansing and Britton, MI with one replicate at each location. Plots were 4 m long, 2-rows wide, with 76 cm row spacing and sown at a rate of 23 seeds m^{-1} row. During 1996, $F_{4:6}$ lines were sown on 16 May in East Lansing and on 23 May in Britton. In 1997, $F_{4:7}$ lines were sown at East Lansing on 4 June and Britton on 10 June.

Plots were rated both years for weight $(100 \text{ seeds})^{-1}$, plant height, maturity date, and lodging. Maturity date was rated as the number of days after 31 August when 95% of the plants in a plot reached their mature pod color (R8) (Fehr et al., 1971). Plant height and lodging where recorded when the plots were mature. Plant height was measured as cm from the ground to the average terminal node of

plants. Plots were rated for lodging on a scale of one to five with one designated as plants standing erect and five as plants lying prostrate to the ground.

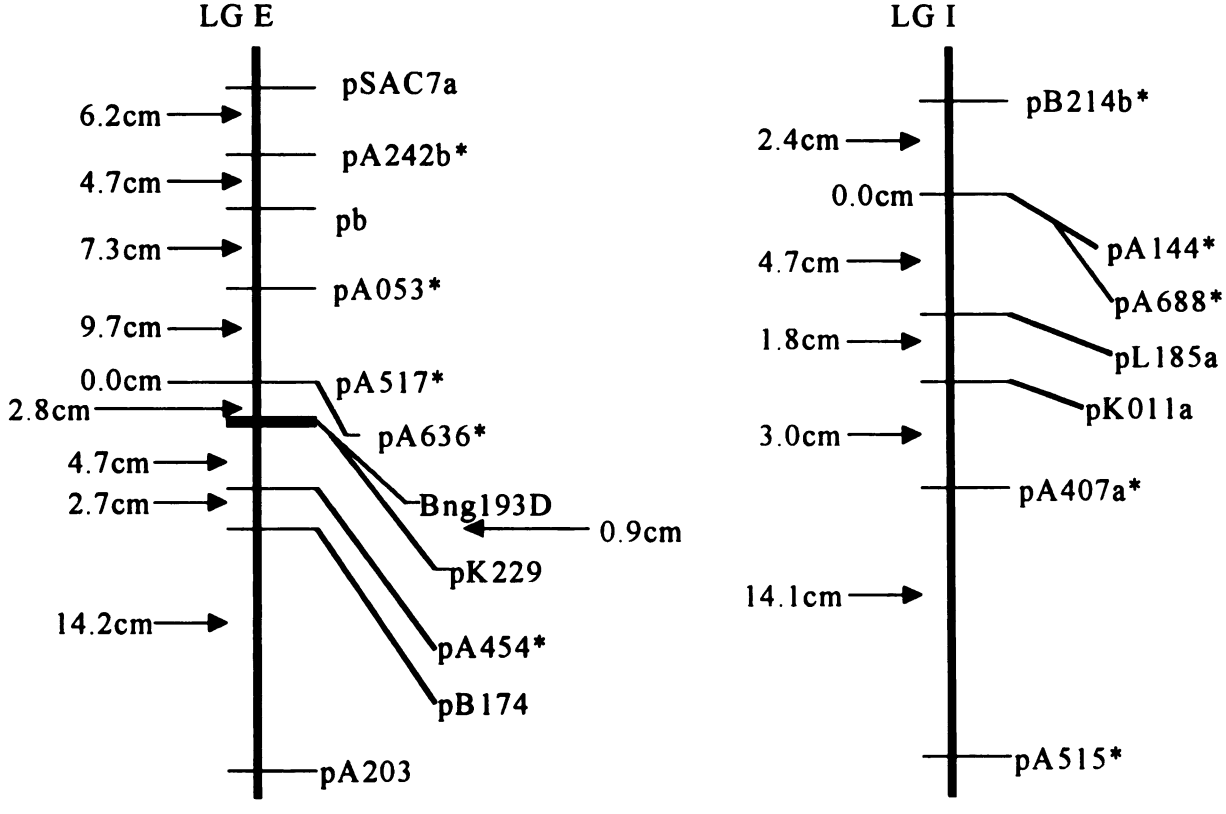
Plots were harvested with a combine to measure seed yield and were not end-trimmed during the growing season. Seed yields were not reported for the E. Lansing location in 1997 because harvest equipment was unable to enter the field due to an unusually wet fall. However, seed was sampled from each plot for protein and oil analyses and weight $(100)^{-1}$ seeds.

Seed protein and oil content was measured at the USDA Northern Regional Research Center at Peoria, IL, by using a Pacific-Scientific NIR grain analyzer. Measurements were taken on a 21 to 25 g sample for each plot. All data collected were analyzed by standard analysis of variance procedures (ANOVA; SAS Institute, 1987). The R^2 value was used to describe the proportion of the genetic variance in the population explained by individual markers in the population.

Results and Discussion

Molecular analyses were conducted using RFLP markers that mapped regions approximately 53 centimorgans (cM) in length on LG E, and approximately 26 cM in length on LG I (Figure 1). Marker loci on LG E were found monomorphic, suggesting that this region was not successfully backcrossed into the population. The RFLP marker B214, which mapped to LG I, was also monomorphic. In Figure 1, the RFLP markers A144 and A688 are shown as completely linked. Results in this study for the two RFLP markers were not

Figure 1: RFLP markers used in this study and their corresponding Linkage Groups (LG).



cm=centimorgans pb=pubescent character

*Markers tested in backcross population (Shoemaker and Specht, 1995)

identical because the mapping population used to generate these linkage groups (Figure 1) was a population other than the interspecific population used in this study.

Only four markers that mapped to LG I were polymorphic. These four RFLP markers were A144, A407, A515, and A688. Genotypic class means for the homozygous *G. soja* and *G. max* classes and the heterozygous class were examined, however, results are only reported for the *G. soja* and *G. max* homozygous class means. The heterozygous class means are not reported due to the small number of heterozygous lines; of the 53 lines in the population, approximately 6% were heterozygous.

All four RFLP markers were significantly ($P < 0.01$) associated with protein content for the combined analyses of both locations for 1996 and 1997 (Table 1). The most significant marker for protein content was A144 (Table 1). This marker explained as high as 76% of the variation for protein content, however, R^2 values were considerably lower in 1997 compared to 1996 data. Oil content was only significant (P 0.05) for both locations in 1996 (Table 2).

Despite lower R^2 values in 1997 for A144, protein and oil means of the genotypic classes for the *G. soja* allele were similar, and in some cases higher, to the 1996 results. For seed protein content, lines ranged from 421 to 485 g (kg seed)$^{-1}$ in 1996 and 420 to 482 g (kg seed)$^{-1}$ in 1997 for the population. For both years of the study, the line with the lowest protein content was still greater than

Table 1. RFLP markers significantly ($P < 0.05$) associated with protein content for 1996 and 1997.

Marker		1996			1997			1996-97
		Britton	E. Lansing	Combined	Britton	E. Lansing	Combined	Combined
A144	$P > F$	0.0001	0.0001	0.0001	0.0003	ns[§]	0.0003	0.0001
	R^2	0.68	0.75	0.76	0.37	0.11	0.34	0.40
					g (kg seed)$^{-1}$			
	MM[†]	427	441	434	437	454	445	436
	SS[‡]	454	472	463	449	463	456	449
A407	$P > F$	0.0001	0.0001	0.0001	0.0002	0.0499	0.0003	0.0001
	R^2	0.66	0.66	0.72	0.27	0.08	0.25	0.28
					g (kg seed)$^{-1}$			
	MM	427	443	435	438	454	446	438
	SS	452	472	462	449	461	455	449
A515	$P > F$	0.0001	0.0001	0.0001	0.0036	ns	0.0072	0.0005
	R^2	0.55	0.62	0.63	0.22	0.04	0.18	0.26
					g (kg seed)$^{-1}$			
	MM	427	443	435	438	455	446	437
	SS	452	472	462	448	460	454	448
A688	$P > F$	0.0001	0.001	0.001	0.0032	0.0113	0.0001	0.0002
	R^2	0.64	0.68	0.70	0.23	0.17	0.33	0.30
					g (kg seed)$^{-1}$			
	MM	426	442	434	438	453	445	437
	SS	453	472	462	447	464	456	447

† MM designates homozygous *G. max* class. ‡ SS designates homozygous *G. soja* class.
§ ns = not significant ($P < 0.05$).

Table 2. RFLP markers significantly ($P < 0.05$) associated with oil content for 1996 and 1997.

Marker		1996			1997			1996-97
		Britton	E. Lansing	Combined	Britton	E. Lansing	Combined	Combined
A144	$P > F$	0.0001	0.0001	0.0001	ns[§]	ns	ns	ns
	R^2	0.64	0.64	0.69	0.05	0.05	0.1	0.05
		g (kg seed)$^{-1}$						
	MM[†]	185	177	181	160	153	156	160
	SS[‡]	169	160	165	156	148	152	156
A407	$P > F$	0.0001	0.0001	0.0001	ns	ns	ns	ns
	R^2	0.57	0.50	0.59	0.02	0.05	0.07	0.02
		g (kg seed)$^{-1}$						
	MM	184	175	180	158	153	158	159
	SS	170	161	165	156	148	152	156
A515	$P > F$	0.0001	0.0001	0.0001	ns	ns	ns	ns
	R^2	0.49	0.38	0.47	0.01	0.04	0.04	0.01
		g(kg seed)$^{-1}$						
	MM	184	174	179	158	152	155	158
	SS	170	162	166	156	150	153	156
A688	$P > F$	0.0001	0.0001	0.0001	ns	ns	ns	ns
	R^2	0.56	0.53	0.60	0.00	0.06	0.08	0.02
		g (kg seed)$^{-1}$						
	MM	185	176	180	158	153	156	159
	SS	169	160	165	157	148	153	157

† MM designates homozygous *G. max* class. ‡ SS designates homozygous *G. soja* class.
§ ns = not significant ($P < 0.05$).

that of the *G. max* parent, which produced 406 g (kg seed)$^{-1}$ in 1996 and 462 g (kg seed)$^{-1}$ in 1997. The *G. soja* parent was previously reported to contain 471 g (kg seed)$^{-1}$ (Diers et al., 1992).

G. soja alleles were associated with greater protein content than *G. max* alleles for all markers significant for protein content on LG I. All *G. max* alleles at significant loci for oil were associated with greater oil content than *G. soja* alleles. This general inverse relationship has been recorded in previous research in several *G. max* populations and *G. max* by *G. soja* populations (Brim and Burton, 1979; Miller and Fehr, 1979; Weber, 1950). *G. soja* marker alleles on LG I were also significantly associated with a decrease in seed yield (Table 3). For the combined analysis across years, the marker A144 revealed that lines homozygous for *G. soja* alleles yielded 174 kg ha^{-1} less than lines homozygous for *G. max* alleles (Table 3).

All markers on LG I were significant ($P < 0.05$) for maturity in 1996 and 1997 (Table 4). At these loci, *G. soja* alleles were associated with earlier maturity than *G. max* alleles. *G. soja* alleles were also associated with smaller seed size (Table 5) when compared to *G. max* alleles. When comparing seed size between the class means for the combined 1996 and 1997 data, *G. soja* alleles were associated with a reduction of seed weight of 0.9 g (100 seed)$^{-1}$.

Table 3. RFLP markers significantly ($P < 0.05$) associated with yield for 1996 and 1997.

Marker		1996			1997	1996-97
		Britton	E. Lansing	Combined	Britton	Combined
A144	$P > F$	0.0001	0.0014	0.0001	ns[§]	0.041
	R^2	0.37	0.29	0.43	0.11	0.15
		kg ha^{-1}				
	MM[†]	2819	3139	2979	2796	2835
	SS[‡]	2493	2829	2661	2661	2661
A407	$P > F$	0.0006	0.0025	0.0001	ns	ns
	R^2	0.23	0.19	0.28	0.03	0.04
		kg ha^{-1}				
	MM	2762	3065	2913	2745	2757
	SS	2515	2848	2681	2670	2670
A515	$P > F$	0.0338	0.0086	0.0041	ns	ns
	R^2	0.13	0.18	0.20	0.03	0.06
		kg ha^{-1}				
	MM	2747	3085	2917	2745	2780
	SS	2551	2833	2692	2659	2659
A688	$P > F$	0.0015	0.0016	0.0001	ns	ns
	R^2	0.25	0.24	0.32	0.07	0.09
		kg ha^{-1}				
	MM	2774	3118	2946	2762	2791
	SS	2510	2827	2669	2655	2655

† MM designates homozygous *G. max* class. ‡ SS designates homozygous *G. soja* class.
§ ns = not significant ($P < 0.05$).

Table 4. RFLP markers significantly ($P < 0.05$) associated with maturity for 1996 and 1997.

Marker		1996			1997			1996-97
		Britton	E. Lansing	Combined	Britton	E. Lansing	Combined	Combined
A144	$P > F$	0.016	0.0005	0.0001	0.0001	0.0003	0.0001	0.0001
	R^2	0.28	0.33	0.38	0.43	0.33	0.44	0.39
	MM†	29	27	28	46	49	47	46
	SS‡	27	24	26	43	45	44	43
A407	$P > F$	0.0002	0.0001	0.0001	0.0001	0.0002	0.0001	0.0001
	R^2	0.26	0.46	0.44	0.37	0.25	0.35	0.38
	MM	29	27	28	46	49	48	46
	SS	27	24	26	43	46	44	43
A515	$P > F$	0.036	0.0001	0.0001	0.0001	0.0002	0.0001	0.0001
	R^2	0.21	0.33	0.33	0.36	0.29	0.38	0.35
	MM	29	27	28	46	49	47	46
	SS	27	24	26	43	45	44	43
A688	$P > F$	0.0002	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001
	R^2	0.31	0.44	0.46	0.39	0.31	0.41	0.39
	MM	29	27	28	46	49	48	46
	SS	27	24	25	43	45	44	43

† MM designates homozygous *G. max* class.

‡ SS designates homozygous *G. soja* class.

Table 5. RFLP markers significantly ($P < 0.05$) associated with weight per 100 seeds for 1996 and 1997.

		1996			1997			1996-97
Marker		Britton	E. Lansing	Combined	Britton	E. Lansing	Combined	Combined
A144	$P > F$	0.0002	0.0018	0.0003	0.0082	0.017	0.0035	0.0052
	R^2	0.35	0.28	0.34	0.24	0.18	0.25	0.23
		g (100 seed)$^{-1}$						
	MM†	15.2	16.1	15.6	14.9	16.8	15.9	14.9
	SS‡	13.6	14.7	14.1	14.0	15.3	14.6	14.0
A407	$P > F$	0.0001	0.0007	0.0001	0.0151	0.0267	0.0082	0.0105
	R^2	0.36	0.23	0.33	0.12	0.10	0.14	0.13
		g (100 seed)$^{-1}$						
	MM	15.2	16	15.6	14.8	16.5	15.6	14.8
	SS	13.7	14.7	14.2	14.1	15.3	14.7	14.1
A515	$P > F$	0.0001	0.009	0.0004	ns§	ns	ns	ns
	R^2	0.34	0.18	0.28	0.05	0.08	0.08	0.06
		g (100 seed)$^{-1}$						
	MM	15.2	15.9	15.6	14.6	16.4	15.6	14.7
	SS	13.6	14.8	14.2	14.6	15.5	14.8	14.2
A688	$P > F$	0.0002	0.0007	0.0002	ns	ns	0.0293	ns
	R^2	0.32	0.27	0.32	0.07	0.12	0.14	0.09
		g (100 seed)$^{-1}$						
	MM	15.2	16.1	15.6	14.7	16.5	15.6	14.8
	SS	13.7	17.8	14.2	14.1	15.1	14.6	14.1

† MM designates homozygous *G. max* class. ‡ SS designates homozygous *G. soja* class.
§ ns = not significant ($P < 0.05$).

Conclusions

Our results show that the QTL on LG I continued to increase protein content after it was backcrossed into a *G. max* background. Furthermore, a QTL, once mapped in an interspecific soybean population, can have a similar genetic effect after it is backcrossed into *G. max*. The effect of the protein QTL in LG I during 1996 was similar to the findings of Diers et al. (1992). However, this QTL had a lesser effect in 1997. This may be due to the late sowing date of the locations in 1997.

Unfortunately, with the increase in protein content, lines with the *G. soja* QTL allele on LG I had lower yield. Because *G. soja* alleles were associated with earlier maturity, this earlier maturity may have resulted in the yield decrease. This may explain the decrease in yield to some extent; however, reasons for this interaction may be more complex. The association between high protein content and low yield may also be caused by the allele for lower yield being tightly linked with the QTL that increased protein content, or caused by the protein QTL directly reducing yield. With more backcrosses to the *G. max* recurrent parent and/or a larger population size, recombination between the gene(s) that decrease yield and the high protein gene may be found. If however, the yield reduction was due to a pleiotrophic effect, then the yield reduction will be impossible to separate from high protein content.

In the study conducted by Diers et al. (1992), QTL for higher protein and oil content were mapped to LG E and I (Shoemaker and Specht, 1995). During the

backcross process, the *pb* gene on LG E was used to select hybrids with the LG E QTL. Although the *pb* gene was segregating in the BC_3 population, the other closely linked markers were not. Because *pb* was not significantly associated with seed protein content in the BC_3 population, recombination between *pb* and the QTL for higher seed protein content likely occurred during backcrossing resulting in the loss of this QTL. Most likely, this was a double crossover event since the RFLP markers that map to LG E surrounding the *pb* gene were monomorphic while the BC_3, F_4-derived lines continued to segregate for this trait.

Further studies should be performed to evaluate the protein QTL in populations other than BC populations, such as in crosses with cultivars. The purpose of these studies would be to determine whether the high protein QTL is effective in other genetic backgrounds and whether the association between low yield and high protein content can be broken.

REFERENCES

Brim, C.A., and J.W. Burton. 1979. Recurrent selection in soybean. II. Selection for increased percent protein in seeds. Crop Sci. 19:494-498.

Brummer E.C., G.L. Graef, J. Orf, J.R. Wilcox, and R.C. Shoemaker. 1997. Mapping QTL for seed protein and oil content in eight soybean populations. Crop Sci. 37:370-378.

Diers, B.W., P. Keim, W.R. Fehr, and R.C. Shoemaker. 1992. RFLP analysis of soybean seed protein and oil content. Theor. Appl. Genet. 83:608-612.

Diers, B.W., and T.C. Osborn. 1994. Genetic diversity of oilseed *Brassica napus* germplasm based on restriction fragment length polymorphisms. Theor. Appl. Genet. 88:662-668.

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

Hartwig, E.E., and K. Hinson. 1972. Association between chemical composition of seed and seed yield of soybean. Crop Sci. 12:829-830.

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

Miller, J.E., and W.R. Fehr. 1979. Direct and indirect selection for protein in soybean. Crop Sci. 19:101-106.

Palmer, R.G., and T.C. Kilen. 1987. Qualitative genetics and cytogenetics. p. 23-48. *In* J.R. Wilcox (ed) Soybean: improvement, production, and uses. 2nd ed. Agron. Monogr. 16. ASA, CSSA, and SSSA, Madison, WI.

SAS Institute. 1987. SAS/STAT guide for personal computers, Version 6 ed. SAS Institute, Cary, NC.

Shoemaker, R.C., and J.E. Specht. 1995. Integration of the soybean molecular and classical genetic linkage groups. Crop Sci. 35:436-446.

Smith, K.J., and W. Huyser. 1987. World distribution and significance of Soybean. p. 1-22. *In* J.R. Wilcox (ed) Soybean: improvement, production, and uses. 2nd ed. Agron. Monogr. 16. ASA, CSSA, and SSSA, Madison, WI.

Weber, C.R. 1950. Inheritance and interaction of some agronomic and chemical characters in an interspecific cross in soybean, *Glycine max* x *G. ussuriencsis.* Ames, Iowa: Research Bulletin 374.

CHAPTER 3

EFFECT OF A HIGH PROTEIN QTL ALLELE FROM *GLYCINE SOJA* IN THREE GENETIC BACKGROUNDS

Introduction

Soybean [*Glycine max*, (L.) Merr.] meal is an important source of protein in many products such as livestock feed, baked goods, and adhesives (Smith and Huyser, 1987). Increasing the protein content of existing breeding lines would prove beneficial and researchers have attempted to increase this (Brim and Burton, 1979; Miller and Fehr, 1979) in *G. max* derived populations. Each of these studies report an increase in seed protein content was associated with a decrease in yield.

A source of genes for increased protein content can be found in *G. soja*, the wild progenitor of *G. max* (Hymowitz and Singh, 1987). Diers et al. (1992) examined a population derived from a cross between the *G. max* experimental line A81-356022 and the *G. soja* plant introduction (PI) 468916 (Diers et al., 1992). Major QTLs from *G. soja* that were significantly associated with an increase in protein content were mapped to linkage groups (LG) E and I using restriction fragment length polymorphism (RFLP) markers. The marker associated with the greatest effect was K011 on LG I, which explained 42% of the variation for protein content in the population.

The two *G. soja* QTLs associated with increased protein content were further analyzed in a backcross population created by selecting one F_2-derived line from the population used by Diers et al. (1992). This line, which carried both QTLs, was a donor parent in three backcrosses, using A81-356022 as the recurrent parent. The new population was developed to determine if the high protein alleles from *G. soja* would have a stable effect in a backcross population. In this backcross population, the QTL allele from *G. soja* on LG I increased protein content (Sebolt and Diers, 1998). Lines with the *G. soja* allele, however, exhibited significantly lower yields than lines without this allele (Sebolt and Diers, 1998). To further evaluate the high protein QTL allele, it was concluded that this allele should be tested in other genetic backgrounds.

The objective of this study was to determine whether the QTL allele from *G. soja* that increases protein content would also increase protein content in crosses with the cultivars 'Parker' and 'Kenwood' and the experimental line C1914. Agronomic traits were analyzed in these populations to determine the effect of the high protein allele on these traits.

Materials and Methods

A population was developed from the interspecific cross between the *G. soja* accession PI 468916 and the *G. max* experimental line A81-356022 (Diers et al., 1992). For our study, one line from the original F_2 population was selected because it was homozygous for the *G. soja* regions associated with increased

protein content on LGs E and I (Diers et al., 1992). This line was used as the donor parent in three backcrosses using A81-356022 as the recurrent parent. During the backcrosses, the *pb* gene on LG E (Palmer and Kilen, 1987), and the RFLP marker A144 on LG I were used to select for the *G. soja* regions where the protein QTL were mapped. A BC_3F_1 plant, that carried both *G. soja* regions, was selfed and the population was inbred to the F_4 generation using single seed descent to develop F_4-derived lines.

One BC_3F_4-derived line from the population was selected, because it was homozygous for the *G. soja* region associated with increased protein content on LG I, and crossed to the cultivars 'Parker' and 'Kenwood' and to the experimental line C1914. The populations were inbred to the F_3 generation using single seed descent. The Parker population included 100 lines, while both the Kenwood and the C1914 populations had 98 lines each.

Leaf tissue was collected from several plants from each $F_{3:4}$ line from the three populations and DNA extractions were conducted according to Kisha et al. (1997) and Southern blotting, hybridization, and autoradiography were performed as described by Diers and Osborn (1994). Four RFLP markers, A144, A407, A515, and A688 were screened against parental line DNA that was digested with four restriction endonucleases (*Eco*RI, *Hind*III, *Dra*I and *Taq*I). The region analyzed was approximately 24 centimorgans (cM) on LG I.

The populations were evaluated as $F_{3:4}$ lines in 1997 and as $F_{3:5}$ lines in 1998. Plots were sown in 1997 on 9 June near East Lansing, MI. Plots were 1 m

long, one-row wide and sown at a rate of approximately 30 seeds m^{-1} row, with a 76 cm row spacing. In 1998, lines were sown in Urbana, IL and near E. Lansing, MI. Lines at the Urbana, Illinois location were sown on 27 April in plots 3.2 m long, two-rows wide, with a 76 cm row spacing and at a rate of 39 seeds m^{-1} row. The E. Lansing location was sown on 12 May in plots 4.3 m long, six-rows wide, with a 38 cm row spacing and at a rate of 25 seeds m^{-1} row. The center four rows in E. Lansing were harvested for yield estimation.

The 1997 and 1998 plots were evaluated for weight $(100 \text{ seeds})^{-1}$ and seed protein and oil contents. In addition, plots were evaluated for plant height, lodging, maturity, and seed yield in 1998. Maturity was rated as the number of days after 31 August when 95% of the plants in a plot reached their mature pod color (R8) (Fehr et al., 1971). Once plants reached maturity, plant height and lodging where recorded. Plant height was measured, in centimeters from the ground to the average terminal node of plants in each plot. Plots were rated for lodging on a scale of one to five with one designated as plants standing erect and five plants lying prostrate to the ground.

All plots were harvested with a combine and seed yield was measured only in 1998. The plots were not end-trimmed during either year. Seed protein and oil content were measured using a Pacific-Scientific NIR grain analyzer at the USDA Northern Regional Research Center in Peoria, IL. Measurements were taken on 21 to 25 g samples.

All data collected were analyzed by standard analysis of variance procedures with PROC GLM of SAS (SAS Institute, 1987). The R^2 value was used to describe the proportion of genetic variance explained by each marker.

Results and Discussion

Molecular analyses were conducted on the populations derived from Parker, Kenwood, and C1914 using the RFLP markers A144, A515, and A688 that mapped to the region approximately 24 centimorgans (cM) in length on LG I (Figure 1). For the three populations studied, marker loci on LG I were polymorphic. Genotypic class means for the homozygous *G. soja* and Parker, Kenwood, and C1914 classes and the heterozygous class were examined, however, results are only reported for the homozygous class means. The heterozygous class means are not reported due to the small number of lines derived from heterozygous plants; approximately 6% were heterozygous for the three populations.

Parker Population:

All three RFLP markers were found to be significant (P · 0.05) (Tables 6 and 7) for seed protein and oil content. The most significant marker for the combined analyses of 1997 and 1998 was A144. This marker explained as much as 44% of the variation for protein content and 15% for oil. R^2 values were mostly consistent across locations and years for protein content. For A144, the lines homozygous for the *G. soja* allele had 20 g (kg seed)$^{-1}$ greater protein content that

Table 6: RFLP markers significantly ($P < 0.05$) associated with protein content for the Parker population.

		A144				A515				A688			
				g (kg seed)$^{-1}$				g (kg seed)$^{-1}$				g (kg seed)$^{-1}$	
Year	Location	$P>F$	R^2	PP†	SS‡	$P>F$	R^2	PP	SS	$P>F$	R^2	PP	SS
1997	E. Lansing	0.0001	0.32	421	441	0.0001	0.17	424	436	0.0001	0.37	421	442
1998	Urbana	0.0001	0.32	428	449	0.003	0.11	432	443	0.0001	0.28	428	447
	E. Lansing	0.0001	0.40	430	449	0.0002	0.16	432	443	0.0001	0.41	429	448
	Combined	0.0001	0.43	429	449	0.0002	0.16	432	443	0.0001	0.40	429	448
1997-98	Combined	0.0001	0.44	426	446	0.0001	0.22	428	441	0.0001	0.43	426	445

† PP designates homozygous 'Parker' class.

‡ SS designates homozygous *G. soja* class.

Table 7: RFLP markers significantly ($P < 0.05$) associated with oil content for the Parker population.

		A144				A515				A688			
				g (kg seed)$^{-1}$				g (kg seed)$^{-1}$				g (kg seed)$^{-1}$	
Year	Location	$P>F$	R^2	PP†	SS‡	$P>F$	R^2	PP	SS	$P>F$	R^2	PP	SS
1997	E. Lansing	ns§	0.01	162	158	ns	0.00	161	160	ns	0.01	162	158
1998	Urbana	0.0002	0.20	193	181	0.03	0.06	191	185	0.0009	0.17	193	183
	E. Lansing	0.005	0.13	187	181	0.003	0.11	187	182	0.01	0.11	188	182
	Combined	0.0001	0.21	190	180	0.004	0.10	189	183	0.0007	0.17	191	182
1997-98	Combined	0.0008	0.15	179	170	0.04	0.05	178	173	0.003	0.13	179	171

† PP designates homozygous 'Parker' class.
‡ SS designates homozygous *G. soja* class.
§ ns = not significant ($P < 0.05$).

lines homozygous for the Parker allele across locations (Table 6). The protein content of Parker was measured as 422 g (kg seed)$^{-1}$ for E. Lansing in 1998 and 400 g (kg seed)$^{-1}$ in Urbana.

The RFLP markers A144 and A688 were found to be significant (P 0.05) for yield in Urbana but not in E. Lansing (Table 8). For A688, the lines homozygous for the *G. soja* allele yielded 251 kg ha^{-1} less than the lines homozygous for the Parker allele in Urbana. Results were based on four harvested rows, which were 4.3 m long, for E. Lansing and two harvested rows, 3.2 m long for Urbana. The Parker parent yielded 3998 kg ha^{-1} in E. Lansing and 1909 kg ha^{-1} in Urbana.

RFLP markers were found to be non significant (P 0.05) for seed weight with the exception of the marker A515 for Urbana (Table 9). Lines homozygous for the *G. soja* allele had a 0.5 g (100 seed)$^{-1}$ less seed weight than the lines homozygous for the Parker allele. RFLP markers were not significant (P 0.05) for maturity, lodging, and height (data not shown).

Kenwood

All three markers, A144, A515, and A688, were significantly (P 0.01) associated with protein and oil content (Tables 10 and 11) for the Kenwood population. The combined analyses of both locations for 1997 and 1998 demonstrated that the RFLP marker A144 explained 41% of the variation for

Table 8: RFLP markers significantly ($P < 0.05$) associated with yield for the Parker population.

		A144				A515				A688			
				kg ha^{-1}				kg ha^{-1}				kg ha^{-1}	
Year	Location	$P>F$	R^2	PP†	SS‡	$P>F$	R^2	PP	SS	$P>F$	R^2	PP	SS
1998	Urbana	0.04	0.08	1806	1615	ns	0.05	1680	1849	0.003	0.15	1827	1576
	E. Lansing	ns§	0.01	4064	3906	ns	0.01	4042	3924	ns	0.02	4049	3850
1998	Combined	ns	0.04	2972	2766	ns	0.02	2983	2829	ns	0.05	2972	2753

† PP designates homozygous 'Parker' class. ‡ SS designates homozygous *G. soja* class.
§ ns = not significant ($P < 0.05$).

Table 9: Genotypic class means for weight per 100 seeds for the Parker population.

		A144				A515				A688			
				g (100 seed)$^{-1}$				g (100 seed)$^{-1}$				g (100 seed)$^{-1}$	
Year	Location	$P>F$	R^2	PP†	SS‡	$P>F$	R^2	PP	SS	$P>F$	R^2	PP	SS
1997	E. Lansing	ns§	0.01	14.9	14.7	ns	0.03	15.1	14.5	ns	0.00	14.8	14.6
1998	Urbana	ns	0.02	14.6	14.1	0.004	0.05	14.7	14.1	ns	0.03	14.6	14.3
	E. Lansing	ns	0.00	18.2	18.4	ns	0.01	18.4	18.1	ns	0.00	18.5	18.2
	Combined	ns	0.00	16.4	16.1	ns	0.03	16.6	16.0	ns	0.02	16.4	16.4
1997-98	Combined	ns	0.01	15.8	15.5	ns	0.04	16.0	15.4	ns	0.02	15.9	15.7

† PP designates homozygous 'Parker' class. ‡ SS designates homozygous *G. soja* class.
§ ns = not significant ($P < 0.05$).

Table 10: RFLP markers significantly ($P < 0.05$) associated with protein content for the Kenwood population.

		A144				A515				A688			
				g (kg seed)$^{-1}$				g (kg seed)$^{-1}$				g (kg seed)$^{-1}$	
Year	Location	$P>F$	R^2	KK†	SS‡	$P>F$	R^2	KK	SS	$P>F$	R^2	KK	SS
1997	E. Lansing	0.0001	0.22	426	441	0.0001	0.27	427	443	0.0001	0.21	426	441
1998	Urbana	0.0001	0.32	430	450	0.0001	0.30	433	450	0.0001	0.28	431	448
	E. Lansing	0.0001	0.50	428	450	0.0001	0.36	432	448	0.0001	0.49	428	448
	Combined	0.0001	0.44	429	450	0.0001	0.36	432	449	0.0001	0.42	430	448
1997-98	Combined	0.0001	0.41	428	447	0.0001	0.39	430	447	0.0001	0.38	428	446

† KK designates homozygous 'Kenwood' class.

‡ SS designates homozygous *G. soja* class.

Table 11: RFLP markers significantly ($P < 0.05$) associated with oil content for the Kenwood population.

		A144				A515				A688			
				g (kg seed)$^{-1}$				g (kg seed)$^{-1}$				g (kg seed)$^{-1}$	
Year	Location	$P>F$	R^2	KK†	SS‡	$P>F$	R^2	KK	SS	$P>F$	R^2	KK	SS
1997	E. Lansing	0.008	0.10	168	162	0.0001	0.18	169	160	0.03	0.09	168	162
1998	Urbana	0.0001	0.26	193	180	0.0001	0.21	190	180	0.0001	0.25	192	180
	E. Lansing	0.0003	0.17	186	179	0.0005	0.13	185	179	0.0005	0.17	186	180
	Combined	0.0001	0.27	189	179	0.0001	0.21	188	179	0.0001	0.26	189	180
1997-98	Combined	0.0001	0.23	182	173	0.0001	0.21	180	173	0.0001	0.22	181	174

† KK designates homozygous 'Kenwood' class.

‡ SS designates homozygous *G. soja* class.

protein content and 23% for oil content in the population. Furthermore, the *G. soja* marker alleles were associated with greater protein content than the Kenwood alleles. Protein content of the lines homozygous for the *G. soja* allele of A144 was 19 g (kg seed)$^{-1}$ greater than the lines homozygous for the Kenwood allele for the combined analysis across 1997 and 1998. The protein content for Kenwood was recorded as 405 g (kg seed)$^{-1}$ for Urbana and 420 g (kg seed)$^{-1}$ for E. Lansing in 1998. Both of these values for Kenwood were lower than the corresponding Kenwood and *G. soja* class means from the population.

All three markers were significant (P 0.05) for yield for Urbana, but none were significant for E. Lansing (Table 12). Lines homozygous for the *G. soja* alleles of A688 had a 138 kg ha^{-1} lower yield compared to lines homozygous for the Kenwood allele (Table 12) in Urbana. In 1998, the Kenwood parent yielded 5013 kg ha^{-1} in E. Lansing and 2029 kg ha^{-1} in Urbana.

The three markers were significant (P 0.05) for seed weight in all locations with the exception of A688 for E. Lansing in 1997 and 1998 (Table 13). The lines homozygous for the *G. soja* allele of A515 had 1 g less (100 seed)$^{-1}$ than the lines homozygous for the Kenwood allele across all three environments. RFLP markers were non significant (P 0.05) for maturity, lodging, and height (data not shown).

Table 12: RFLP markers significantly ($P < 0.05$) associated with yield for the Kenwood population.

		A144				A515				A688			
				kg ha^{-1}				kg ha^{-1}				kg ha^{-1}	
Year	Location	$P>F$	R^2	KK†	SS‡	$P>F$	R^2	KK	SS	$P>F$	R^2	KK	SS
1998	Urbana	0.04	0.07	1797	1678	0.03	0.05	1790	1681	0.03	0.08	1789	1651
	E. Lansing	ns§	0.04	4295	4101	ns	0.01	4232	4142	ns	0.03	4270	4135
1998	Combined	ns	0.06	3046	2890	ns	0.03	3011	2912	ns	.06	3030	2893

† KK designates homozygous 'Kenwood' class.

‡ SS designates homozygous *G. soja* class.

§ ns = not significant ($P < 0.05$).

Table 13: Genotypic class means for weight per 100 seeds for the Kenwood population.

		A144				A515				A688			
				g (100 seed)$^{-1}$				g (100 seed)$^{-1}$				g (100 seed)$^{-1}$	
Year	Location	$P>F$	R^2	KK†	SS‡	$P>F$	R^2	KK	SS	$P>F$	R^2	KK	SS
1997	E. Lansing	0.005	0.11	15.1	14.1	0.001	0.11	15.2	14.2	ns§	0.06	15.0	14.2
1998	Urbana	0.0001	0.19	13.9	12.6	0.0001	0.24	13.9	12.6	0.0006	0.17	13.7	12.7
	E. Lansing	0.04	0.07	17.8	16.8	0.006	0.08	17.9	16.9	ns	0.02	17.6	17.2
	Combined	0.002	0.13	15.8	14.7	0.0001	0.16	15.9	14.8	ns	0.07	15.7	14.9
1997-98	Combined	0.006	0.10	15.5	14.5	0.0007	0.12	15.5	14.6	ns	0.06	15.4	14.7

† KK designates homozygous 'Kenwood' class. ‡ SS designates homozygous *G. soja* class.
§ ns = not significant ($P < 0.05$).

C1914

The three RFLP markers were not significant (P 0.05) for seed protein and oil content at any location or over locations for the C1914 population (Table 14). The *G. soja* genotypic class means for protein and oil content were similar to the experimental line C1914. The means of both were 472 g (kg seed)$^{-1}$ for the combined analyses for 1997 and 1998 for all three markers. Furthermore, oil contents for all three markers were also similar when the C1914 alleles were compared to *G. soja* alleles. Yield was not significantly different (P 0.05) at any location or across locations (Table 15), however, a 57 kg ha^{-1} reduction in yield was found when *G. soja* genotypic class means were compared with the C1914 class means for Urbana for the marker A144. A 202 kg ha^{-1} reduction in yield of the *G. soja* alleles for E. Lansing in 1998 was observed.

Table 14: Protein and oil content means of the genotypic classes in the C1914 population.

		Protein						Oil					
		A144		A515		A688		A144		A515		A688	
		g (kg seed)$^{-1}$											
Year	Location	CC†	SS‡	CC	SS	CC	SS	CC	SS	CC	SS	CC	SS
1997	E. Lansing	466	463	464	465	466	463	150	151	150	150	150	151
1998	Urbana	480	482	482	481	480	482	160	160	160	160	160	162
	E. Lansing	470	471	470	470	470	470	167	168	167	168	167	168
	Combined	475	477	475	476	475	476	164	164	163	164	164	165
1997-98	Combined	472	472	472	472	472	472	161	159	159	159	159	161

† CC designates homozygous 'C1914' class. ‡ SS designates homozygous *G. soja* class.

Table 15: Yield means of the genotypic classes in the C1914 population.

		A144		A515		A688	
		kg ha^{-1}		kg ha^{-1}		kg ha^{-1}	
Year	Location	CC†	SS‡	CC	SS	CC	SS
1998	Urbana	1660	1603	1667	1589	1658	1578
	E. Lansing	3851	3649	3852	3683	3844	3649

† CC designates homozygous 'C1914' class.
‡ SS designates homozygous *G. soja* class.

Conclusions

The objective of this study was to test if a QTL allele from *G. soja* would increase protein content in three different genetic backgrounds. This *G. soja* allele increased protein content in the Kenwood and Parker backgrounds, but not in the C1914 background. This suggests that there is a high protein allele that is allelic with the *G. soja* allele in C1914 but not in Parker or Kenwood. Seed yields in E. Lansing were twice as great as in the Urbana location. The low yields in Urbana were at least partly the result of the populations being grown outside their range of adaptation and that seed filling occurred during a dry period.

The high protein allele from *G. soja* was associated with less yield in crosses with Kenwood and Parker but not in C1914. The data provide additional evidence that the allele that increases protein content also lowers yield. If the high protein gene was associated with less yield due to a coupling linkage with a yield reducing gene, this coupling linkage would have to be present for both the gene from *G. soja* and C1914. This is less likely than the high protein gene causing the yield reduction.

If there was a coupling linkage between the genes that contributed to decrease seed yields and the allele for higher protein content, eventually, through recombination, these effects could be separated. To do this, a larger population size would need to be studied and/or the number of backcrosses to the recurrent parent would have to be increased. If it were a pleiotrophic effect, then these traits would never be separated.

In nearly all studies conducted to increase seed protein content, yield reductions were found when protein content was increased. Not only should the question of pleiotrophy be considered when deciding whether to continue with this research, but also calculations for protein per hectare. Despite the significant increase in seed protein content, protein per hectare was only marginally greater in lines homozygous for the *G. soja* QTL for increased protein content because of the associated decrease in seed yield. For example, lines homozygous for the *G. soja* allele for A144 produced 1754 kg ha^{-1} protein while lines homozygous for the Parker allele produced 1748 kg ha^{-1} protein in E. Lansing during 1998. Unquestionably, because protein per hectare for the Parker parent was only slightly lower than lines associated with the *G. soja* homozygous class, it would not be advantageous to continue research if the high protein gene caused the yield reduction.

REFERENCES

Brim, C.A., and J.W. Burton. 1979. Recurrent selection in soybean. II. Selection for increased percent protein in seeds. Crop Sci. 19:494-498.

Diers, B.W., P. Keim, W.R. Fehr, and R.C. Shoemaker. 1992. RFLP analysis of soybean seed protein and oil content. Theor. Appl. Genet. 83:608-612.

Diers, B.W., and T.C. Osborn. 1994. Genetic diversity of oilseed *Brassica napus* germplasm based on restriction fragment length polymorphisms. Theor. Appl. Genet. 88:662-668.

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

Hymowitz, T., and R.J. Singh. 1987. Taxonomy and speciation. In: Wilcox JR (ed) Soybean: improvement, productions, and uses, 2nd edn. Agronomy 16:23-48.

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

Miller, J.E., and W.R. Fehr. 1979. Direct and indirect selection for protein in soybean. Crop Sci. 19:101-106.

Palmer, R.G., and T.C. Kilen. 1987. Qualitative genetics and cytogenetics. p. 23-48. *In* J.R. Wilcox (ed) Soybean: improvement, production, and uses. 2nd ed. Agron. Monogr. 16. ASA, CSSA, and SSSA, Madison, WI.

SAS Institute. 1987. SAS/STAT guide for personal computers, Version 6 ed. SAS Institute, Cary, NC.

Sebolt, A.M., and B.W. Diers. 1998. Evaluation of genes from the wild *Glycine soja* that increase protein content in soybean. p. 69. *In* 1998 Agronomy abstracts. ASA, Madison, WI.

Smith, K.J., and W. Huyser. 1987. World distribution and significance of soybean. p. 1-22. *In* J.R. Wilcox (ed) Soybean: improvement, productions, and uses. 2nd edn. Agron. Monogr. 16. ASA, CSSA, and SSSA, Madison, WI.