

To the Graduate Council:

I am submitting herewith a dissertation written by Dilip Raj Panthee entitled "Genetic Mapping of Quantitative Trait Loci Conditioning Protein Concentration and Quality, and Other Seed Characteristics in Soybean [*Glycine max* (L.) Merrill]". I have examined the final electronic copy of this dissertation for form and content and recommend that it be accepted in partial fulfillment of the requirements for the degree of Doctor of Philosophy, with a major in Plants, Soils, and Insects.

Vincent R. Pantalone

Major Professor

We have read this dissertation
and recommend its acceptance:

Arnold M. Saxton

Dennis R. West

Carl E. Sams

Accepted for the Council:

Anne Mayhew

Vice Chancellor and
Dean of Graduate Studies

(Original signatures are on file with official student records.)

**GENETIC MAPPING OF QUANTITATIVE TRAIT LOCI CONDITIONING
PROTEIN CONCENTRATION AND QUALITY, AND OTHER SEED
CHARACTERISTICS IN SOYBEAN [*GLYCINE MAX* (L.) MERRILL]**

A Dissertation

Presented for the

Doctor of Philosophy

Degree

The University of Tennessee, Knoxville

Dilip Raj Panthee

May 2005

Dedicated to
My Respected Mother



Mrs. Dharma Kumari Panthee
for Her Continuous Inspiration and Support
for My Education
She Provided Me in My Life

ACKNOWLEDGEMENT

I am thankful to all those who provided their genuine help to complete my PhD research and bring the dissertation to this stage. I am most grateful to Dr. Vincent R. Pantalone, my major professor, for his generous support, encouragement and guidance throughout my PhD program. Committee members Dr. Arnold M. Saxton, Dr. Carl E. Sams and Dr. Dennis R. West are thanked for their valuable guidance and suggestions at various stages of my research and preparation of dissertation. Members of the soybean breeding team Dr. Madeleine Spencer, Deborah Landau-Ellis, Elizabeth Myer, David, Walker, Andrew Scaboo and others who worked in various seasons are thanked for their help in completing the field and laboratory experiment. Special thanks go to the United Soybean Board and Tennessee Soybean Promotion Board for their financial support.

I am thankful to the Institute of Agriculture and Animal Sciences, Tribhuvan University, Nepal for providing me a long study leave to pursue my PhD. Support of Sarita (my wife), Jyotsna (daughter) and Jwalanta (son) in all aspects of my study and family life is appreciated. Finally, all who helped to make my stay comfortable and joyful at Knoxville are also thanked.

ABSTRACT

Soybean [*Glycine max* (L.) Merr.] is grown for its high oil and protein concentration, which may be used in a wide array of industries throughout the world. It contributes about 15 billion dollars annually to the US economy. Focusing on protein concentration and quality, I was interested to discover genomic regions as quantitative trait loci (QTL) associated with nitrogen accumulation during reproductive stages, protein concentration, storage protein fractions, amino acid composition, seed size, oil content, and agronomic traits. A population of 101 F₆-derived recombinant inbred lines (RIL) created from a cross of N87-984-16 x TN93-99 were used to achieve these objectives. Experiments were conducted in six environments during 2002 and 2003. A significant ($P < 0.05$) difference among genotypes was found for all agronomic and seed composition traits. A total of 94 polymorphic simple sequence repeat (SSR) molecular genetic markers were screened in DNA from progenies. Single factor ANOVA in SAS software was used to identify candidate QTL that were confirmed by composite interval mapping using QTL Cartographer. QTL governing nitrogen accumulation in soybean seed were distributed among seven linkage groups. Two QTL for protein concentration, seven QTL for oil concentration and eight QTL for seed size were identified. Moreover, three QTL for glycinin (near Satt461, Satt292 and Satt156) and two QTL for conglycinin (near Satt461 and Satt249) were detected in this

population. Nine QTL (near Satt002, Satt185, Satt201, Satt235, Satt252, Satt268, Satt427, Satt436 and Satt590) distributed on six molecular linkage groups were associated with cysteine (Cys) and four QTL (near Satt002, Satt252, Satt564 and Satt590) distributed on four molecular linkage groups were associated with methionine (Met) concentration in soybean seed. At least one QTL for each amino acid was detected in this population. Two QTL each for lodging, plant height and seed filling period, six for maturity and three for yield were detected in this population. Three lines developed from this population were released as germplasm by the Tennessee Agricultural Experiment Station. TN03-350 and TN04-5321 were released because those lines combine high yield and increased protein concentration. TN04-5363 was released because it combines high yield and total sulfur containing amino acids (Cys and Met). The QTL identified and the germplasms developed from this study may be useful in improving the nutritional quality of soybean.

TABLE OF CONTENT

PART 1	1
INTRODUCTION	1
References	12
PART 2	15
Genomic Regions Governing Soybean Seed Nitrogen Accumulation	15
Abstract	16
Introduction	17
Materials and Methods	19
Results and Discussion	23
References	30
PART 3	34
Quantitative Trait Loci for Seed Protein and Oil Concentration, and Seed Size in Soybean	34
Abstract	35
Introduction	37
Materials and Methods	40
Results and Discussion	45
References	62
PART 4	68
Quantitative Trait Loci for β-Conglycinin (7S) and Glycinin (11S) Fractions of Soybean Storage Protein	68
Abstract	69
Introduction	70
Materials and Methods	73
Results and Discussion	80
References	94
PART 5	99
Quantitative Trait Loci Controlling Sulfur Containing Amino Acids, Methionine and Cysteine, in Soybean Seeds	99
Abstract	100
Introduction	101
Materials and Methods	106
Results and Discussion	111
References	124
PART 6	132
Genomic Regions Associated with Amino Acid Composition in Soybean	132
Abstract	133
Introduction	135
Materials and Methods	138

Results and Discussion-----	142
References -----	163
PART 7	169
Quantitative Trait Loci for Agronomic Traits in Soybean	169
Abstract-----	170
Introduction -----	171
Materials and Methods -----	174
Results and Discussion-----	180
References -----	195
PART 8	202
Registration of Improved Soybean Protein Germplasms TN03-350,	
TN04-5321 and TN04-5363	202
References -----	208
PART 9	210
CONCLUSION AND FUTURE RESEARCH.....	210
VITA	214

LIST OF TABLES

Table 1.1: Estimated area, yield and production of soybean in world, 2004.	6
Table 1.2: Production of soybean and soybean-based items (million metric tones) and trade in different countries.	7
Table 2.1: Descriptive statistics for nitrogen and protein content on a dry weight basis in a soybean RIL population (n=105) measured at three reproductive growth stages.	24
Table 2.2: Molecular markers and molecular linkage groups for QTL governing nitrogen accumulation in developing seeds of soybean at three reproductive growth stages.	27
Table 2.3: Soybean seed nitrogen accumulation QTL at three reproductive growth stages, which were also, detected as amino acid QTL.	29
Table 3.1: Descriptive statistics for seed protein and oil concentration, and for seed size in an F6-derived soybean population of N87-984-16 x TN93-99 averaged over six environments (2002 and 2003).	46
Table 3.2: Genetic and phenotypic correlations among protein, oil and seed size in F6-derived soybean population from N87-984-16 x TN93-99 averaged over six environments (2002 and 2003).	49
Table 3.3: Quantitative trait loci, map position and genetic contribution for protein, oil and seed size in an F6-derived soybean population from N87-984-16 x TN93-99.	51
Table 4.1: Descriptive statistics of soy protein fraction from a population of 101 F6 – derived RIL of N87-984-16 x TN93-99 (average of 2001 and 2003).	81
Table 4.2: Correlation coefficient between conglycinin, glycinin and their ratio and individual components on the basis of combined least square means from 2001 and 2003.	82
Table 4.3: QTL detected for soybean storage protein fractions by composite interval mapping analysis in F6-derived population of N87-984-16 x TN93-99.	87
Table 5.1: Descriptive statistics (seed dry weight basis) for cysteine (Cys), methionine (Met), total sulfur containing amino acids (Met+Cys) and total seed protein concentration averaged over six environments in a soybean RIL population derived from N87-984-16 x TN93-99.	112
Table 5.2: Genetic and phenotypic correlation coefficient among Cys, Met and other traits in an F6-derived RIL population of soybean from N87-984-16 x TN93-99.	116
Table 5.3: QTL detected by composite interval mapping for cysteine, methionine and Cys + Met in F6-derived RIL population of soybean from N87-984-16 x TN93-99.	118
Table 6.1: Descriptive statistics (g kg ⁻¹ seed dry weight) of various amino acid and total protein concentration in soybean seed in an F6-derived population of N87-984-16 x TN93-99.	143
Table 6.2: Ratio between essential to non-essential (E:NE), and essential to total (E:T) amino acids in an F6-derived population of soybean from N87-984-16 x TN93-99.	147
Table 6.3: Simple phenotypic correlation coefficients between amino acids in soybean seeds in an F6-derived population of N87-984-16 x TN93-99.	149
Table 6.4: Quantitative trait loci associated with essential amino acids in soybean seed in an F6-derived population of N87-984-16 x TN93-99.	151
Table 6.5: Quantitative trait loci associated with non-essential amino acids in soybean seed in an F6-derived population of N87-984-16 x TN93-99.	156
Table 6.6: QTL† associated with various amino acids in soybean F6-derived RIL population from N87-984-16 x TN93-99.	160
Table 6.7: Linkage group and QTL† controlling various amino acids in a soybean F6-derived RIL population from N87-984-16 x TN93-99.	162
Table 7.1: Descriptive statistics for agronomic traits in an F6-derived soybean population developed from N87-984-16 x TN93-99.	180

Table 7.2: Genetic and phenotypic correlation coefficients among agronomic traits in an F6-derived soybean population developed from N87-984-16 x TN93-99.	183
Table 7.3: Quantitative trait loci, their position and genetic contribution as additive effects for agronomic traits in an F6-derived soybean population developed from N87-984-16 x TN93-99.	185
Table 7.4: Number of QTL in various linkage groups associated with agronomic traits in an F6-derived soybean population developed from N87-984-16 x TN93-99.	193

LIST OF FIGURES

Figure 2.1: Frequency distribution for a population of soybean recombinant inbred lines for seed nitrogen content class means (constructed at intervals of 0.2% nitrogen), averaged over three reproductive growth stages.....	24
Figure 3.1: Frequency distribution of seed protein concentration (g kg ⁻¹) in an F6-derived soybean population of N87-984-16 x TN93-99 averaged over six environments (2002 and 2003).....	46
Figure 3.2: Frequency distribution of seed oil concentration (g kg ⁻¹) in an F6-derived soybean population of N87-984-16 x TN93-99 averaged over six environments (2002 and 2003).....	47
Figure 3.3: Frequency distribution of seed size (mg seed ⁻¹) in an F6-derived soybean population of N87-984-16 x TN93-99 averaged over six environments (2002 and 2003).....	48
Figure 3.4: LOD score plot for protein concentration on LG G in combined data. There is a major QTL on LG G near Satt570.....	52
Figure 3.5: LOD score plot for protein concentration on LG D1b in combined data averaged over six environments (2002 and 2003). There is a major QTL on LG D1b Satt274.	53
Figure 3.6: LOD score plot for seed size on LG D1a in combined data averaged over six environments (2002 and 2003). There is a major QTL flanked by Satt147 and Satt184 on LG D1a.	59
Figure 4.1: Banding pattern of the components of 7S (a', a and b) and 11S (acidic and basic) fraction of soy protein in SDS-PAGE obtained from soybean RIL of N87-984-16 x TN93-99.	78
Figure 4.2: Frequency distribution of glycinin in a population of F6 –derived RIL of N87-984-16 x TN93-99.	83
Figure 4.3: Frequency distribution of b-conglycinin in a population of F6 –derived RIL of N87-984-16 x TN93-99.	83
Figure 4.4: Scatter plot of glycinin vs. b-conglycinin in RIL population developed from N87-984-16 x TN93-99. High level of glycinin is present irrespective of b-conglycinin content.	84
Figure 4.5: LOD score plot for glycinin on LG G. A major QTL appears located between Satt564 and Satt191, and another major QTL appears located downstream of Satt191.....	90
Figure 5.1: Distribution of Met (g kg ⁻¹ seed dry matter) in an F6-derived RIL population of soybean developed from a cross of N87-984-16 x TN93-99.....	112
Figure 5.2: Distribution of Cys (g kg ⁻¹ seed dry matter) in an F6-derived RIL population of soybean developed from a cross of N87-984-16 x TN93-99.....	113
Figure 5.3: Distribution of Met+Cys (g kg ⁻¹ seed dry matter) in an F6- derived RIL population of soybean developed from a cross of N87-984-16 x TN93-99.	113
Figure 5.4: Scatter diagram of total S-containing amino acid (% of protein) in a soybean RIL population derived from N87-984-16 x TN93-99.....	117
Figure 5.5: LOD score plot on LG G for Methionine QTL. Likely position of Met QTL is upstream of Satt564 on LG G.....	121
Figure 7.1: Frequency distribution for plant height in an F6-derived soybean population developed from N87-984-16 x TN93-99.	181
Figure 7.2: Frequency distribution for maturity in an F6-derived soybean population developed from N87-984-16 x TN93-99.	181
Figure 7.3: Frequency distribution for yield in an F6-derived soybean population developed from N87-984-16 x TN93-99.	182
Figure 7.4: LOD score plot for seed filling period on LG A1 in an F6-derived soybean population developed from N87-984-16 x TN93-99.	186

Figure 7.5: LOD score plot for lodging on LG A1 in an F6-derived soybean population developed from N87-984-16 x TN93-99.	188
Figure 7.6: LOD score plot for maturity on LG A1 in an F6-derived soybean population developed from N87-984-16 x TN93-99.	189
Figure 7.7: LOD score plot for seed yield on LG A1 in an F6-derived soybean population developed from N87-984-16 x TN93-99.	191
Figure 7.8: LOD score plot for seed yield on LG D1b in an F6-derived soybean population developed from N87-984-16 x TN93-99.	191

PART 1

INTRODUCTION

The soybean [*Glycine max* (L.) Merrill] ($2n = 2X = 40$) is an important member of the family *Leguminosae* and sub-family of *Papilionoideae*. It is an annual herb mainly grown for seed from which oil and protein are extracted. Soybean cultivars typically achieve 75 to 125 cm in height with sparse or dense branches depending on cultivar and cultivation condition. There are four different types of leaves: the first pair of simple cotyledons or seed leaves, the second pair of simple primary leaves, trifoliate foliage leaves, and the prophylls (Lersten and Carlson, 2004). Soybean has a tap root system with diffused roots bearing several root nodules. The first root nodule appears about 10 days after planting and nodule formation continues throughout vegetative and reproductive phases. The nodules supply the plants primary nitrogen requirements.

The growth habit of soybean can be determinate or indeterminate. In the indeterminate type of soybean, the inflorescences are auxiliary racemes and plants at maturity have a sparse and rather even distribution of pods on all branches with a diminishing frequency towards the tip of the stem. In the determinate type of plants, both an auxiliary raceme and a terminal raceme with a cluster of flowers are present. The flowers are in a cluster of 2 to 35 (Carlson and Lersten, 2004) but only a few (about 20%) develop into pods. Generally, cultivars adapted to northern regions are indeterminate and those towards southern regions are determinate. However, farmers' desire to harvest the crop earlier has led to an influx of

more indeterminate cultivars in the Southern region in recent years. Soybean has a papilionaceous complete flower with a tubular calyx of five unequal sepal lobes and a five parted corolla consisting of a posterior banner petal, two lateral wing petals, and two anterior keel petals in contact with each other but not fused. The ten stamens occur in two groups in which the filaments of nine of the stamens are fused and elevated as a single structure, whereas the posterior stamen remains separate. The single pistil is unicarpellate and has one to four ovules alternating along the posterior suture.

Soybean is believed to have been originated from China (Hymowitz and Singh, 1987). It is believed that soybean was domesticated in China about 5000 years ago. From the site of domestication, it was spread to South China, Korea, Japan and South-East Asian countries. It was introduced into the USA sometime in the 17th century. In the beginning, it was used as a forage crop which continued through the 1930s (Poehlman and Sleper, 1995). Thereafter people started to use soybean as a seed crop and plant improvement work for seed yield and quality was started. Now soybean is one of the top priority crops in research because of its food, livestock feed, and industrial applications.

The genus *Glycine* is composed of two subgenera *Glycine* and *Soja*. The wild perennial soybean belongs to the subgenus *Glycine* whereas

cultivated soybean (*Glycine max*) and its wild annual progenitor (*Glycine soja*) belong to the subgenus *Soja*. There are more than 100,000 accessions of *Glycine max* and about 10,000 accessions of *Glycine soja* in the world (Palmer et al., 1996). There is a good deal of variation in this large genetic pool, which is being utilized for the improvement of characters of economic, nutritional and industrial importance.

In addition to *G. max*, the following species are important sources of genetic variation: *G. clandestina* Wendl, *G. clandestina* var. *sericea*, *G. falcata* Benth, *G. latrobeana* (meissn) Benth, *G. tabacina* (Labill) Benth., *G. tabacina* var. *latifolia*, *G. tabacina* var. *uncinata*, *G. tomentella* Hayata, and *G. sericea* Benth (Hymowitz and Singh, 1987). Some of the successful interspecies hybrids are (*G. falcata* x *G. clandestine*), (*G. canescenes* x *G. clandestine*), (*G. falcata* x *G. tomentella*), (*G. latifolia* x *G. tabacina*), (*G. latifolia* x *G. tomentella*), (*G. latrobeana* x *G. tabacina*), (*G. latrobeana* x *G. canescenes*), and (*G. tomentella* x *G. tabacina*) (Hymowitz and Singh, 1987). Of these species, *G. max*, *G. soja* and *G. ussuriensis* are diploid ($2n = 2X = 40$) and cross compatible (Palmer et al., 1996). However, in some of the species, the basic chromosome number has been reported as 10 (Palmer and Kilen, 1987). Using this genetic information and on the basis of crossing compatibility, crop improvement efforts were made in the past by traditional breeding methods. This

brought about significant improvements particularly in yield, disease resistance, seed quality and maturity.

Soybean is one of the oldest crops with a range of uses in human life. Although it was used as a fodder until the 1930s in the USA, seed production is currently the main focus. Soybean has long been recognized as a valuable component of medicine, food and feed in various parts of the world including China (Mounts et al., 1987). Soybean oil and protein are mainly used for animal feed and increased attention is now given to its use as human food. Some of the food types that are part of life in Asia are soy milk, tofu, soy sprouts, immature boiled soybeans, roasted soybeans, soy sauce, soy paste, miso, natto, tempeh, and soy nuggets. Liu (1997) has described in detail the available methods for preparation of these food items and their utilization. From his descriptions it is clear that these items were popular for hundreds of years in China and East Asia. As the biochemistry of soybean became better understood, the role of various seed components became more evident in determining the quality of food items. For example, various fractions of soy proteins play an important role in determining the quality of tofu; seed size determines sprout quality; lipoxygenase concentration determines the beany flavor of soy milk. Such understandings are necessary to improve the overall quality of soybean. Furthermore, there is currently great interest in utilizing isoflavones, which are available in soybean, as cancer treatment. Industrial uses of soybean

oil include bio-diesel, inks, lubricants, nylons, plasticizers, coatings, and cosmetics. Thus, it is clear that soybean continues to have multiple uses of commercial interest.

Because of the importance of the crop, along with improvements in genetics, agricultural engineering, pest management and agronomic practices, the area under soybean production and the total production increased drastically over time. Now it is estimated that about 194.6 million metric tonnes (mt) of soybean is produced as an oil seed crop annually, which is almost 60% of the total oil seed production in the world (Table 1.1 and 1.2). In terms of GNP value, it is estimated at \$15 billion in USA and \$28.3 billion in the world. Soybean has occupied first place in protein and oil production among the oil seed crops.

Table 1.1: Estimated area, yield and production of soybean in world, 2004.

Country	Area (Million ha)	Yield (mt/ha)	Production (Million mt)
USA	29.20	2.54	74.29
Brazil	18.40	2.83	52.00
China	9.40	1.76	16.51
Argentina	12.50	2.80	35.00
European Union	0.30	3.21	0.95
PARAGUAY	1.45	2.69	3.90
Others	10.54	1.13	11.95
Total	81.79	2.38	194.60

Source:(Soytech Inc., 2004)

Table 1.2: Production of soybean and soybean-based items (million metric tones) and trade in different countries.

Description	Item	Quantity
World oilseed production	Soybean	194.6
	Cottonseed	33.00
	Peanut	30.73
	Sunflower	23.79
	Rapeseed	31.87
	Others	12.77
Total		326.76
World oilseed export	Soybean	63.52
	Cottonseed	1.15
	Peanut	1.45
	Sunflower	2.95
	Rapeseed	6.08
	Others	0.29
Total		75.43
World vegetable oil production	Soybean	30.61
	Palm	26.24
	Sunflower seed	8.26
	Rapeseed	11.51
	Cotton seed	3.53
	Peanut	4.30
	Coconut	3.19
	Olive	2.35
	Palm kernel	3.30
Total		93.28
World vegetable protein meal	Soybean	131.95
	Cotton seed	11.29
	Rapeseed	18.48
	Sunflower seed	9.05
	Fish	5.76
	Peanut	5.43
	Copra	1.67
	Palm kernel	3.93
Total		187.56

Source: (Soytech Inc., 2004)

Considering its importance in the past 60 years, soybean improvement has received a high priority and hundreds of cultivars have been developed. The main characteristics improved by conventional breeding methods are yield, maturity and disease resistance. However, several important characters like seed quality, and yield are controlled by polygenes¹, which cannot be easily improved by conventional breeding. Molecular techniques, beginning in the 1980s, have become useful tools in crop improvement programs. Before embarking on the aspects of using molecular techniques in crop improvement programs, it is worth considering the genome composition of soybean.

It is estimated that soybean has 1.29×10^9 to 1.81×10^9 base pairs in haploid DNA content (Shoemaker et al., 1996). More than 35% of the genome is heterochromatic and approximately 40-60% of the genome is composed of repetitive sequences. About 65-70% of single copy sequences are estimated to have a short period interspersion with single copy sequences of 1.1-1.4 kbp alternating with repetitive sequence elements of 0.3-0.4 kb (Shoemaker et al., 1996). Currently, it is estimated that there are more than 200 genes for simple qualitative characters in soybean (Poehlman and Sleper, 1995). There are many minor genes controlling a character of economic interest, particularly yield and seed quality traits. These are quantitative traits, which show continuous

¹ Two or more genes at different loci, which produce cumulative, effects on a single genetic trait and shows continuous variation.

variation and it is difficult to make selections by conventional methods. Genes are located at a particular location on a chromosome, called a locus. Quantitative traits are controlled by many regions on the chromosomes, but each such region may have multiple genes, or regulatory elements, simply referred to as quantitative trait loci (QTL). Such QTL can be identified if there are markers associated with them. To identify a QTL, it is important to map the loci controlling the trait of interest on the chromosome, then identify the markers that are associated with the trait. Once these tasks are done, the markers can be used in selection to improve the trait of interest.

Major QTL mapped in the soybean genome for various traits have included seed yield, maturity, plant height, lodging, hard seededness, seed protein, seed oil, seed weight, and cyst nematode resistance (Shoemaker et al., 1996). With the advent of various mapping techniques, a high-density linkage map of soybean has been developed. The present composite molecular map consists of more than 1850 molecular markers distributed among 20 linkage groups spanning about 2500 cM (Cregan et al., 1999; Song et al., 2004). This map is based on simple sequence repeat (SSR), restriction fragment length polymorphism (RFLP), random amplified polymorphic DNA (RAPD), amplified fragment length polymorphism (AFLP) molecular markers and protein isozymes.

RFLP are robust for constructing a map, the technique produces only two parental allele types. Thus the polymorphic potential of an RFLP probe is primarily a function of the frequency of the two alleles at each locus that the probe detects. In contrast, simple sequence repeat markers are single locus markers with multiple alleles serving as co-dominant markers (Cregan et al., 1999). Therefore, a DNA marker capable of detecting the polymorphism even in the event of multiple alleles at a single locus would be more useful. SSR markers are dependent on the number of alleles and their frequencies to determine the polymorphism (Cregan et al., 1999). As mentioned before, soybean is believed to be an ancient polyploid in the course of its evolution, hence it has several repetitive DNA sequences in the genome. Hence SSR markers are more useful for mapping species like soybean. Considering the limitations of RFLP and the genome composition of soybean, researchers have developed more than 1000 SSR markers (Cregan et al., 1999; Song et al., 2004), which are publicly available.

An SSR or microsatellite is a small segment of DNA, usually 2 to 5 bp in length that repeats itself a number of times. Useful SSRs usually repeat the core motif 9-30 times. Some of the major core motifs that are used in the development of SSR markers for soybean include ATT, AT, CTT, and CT. In general, loci with ATT core motif are most informative followed by those with AT cores (Cregan et al., 1999). The regions flanking the

microsatellite are generally conserved among genotypes of the same species and polymerase chain reaction (PCR) primers to the flanking regions can be used to amplify SSR DNA fragments (Cregan et al., 1999). Length polymorphisms are created when PCR products from different alleles vary in length as a result of variation in the number of repeat units in the SSR. These can be analyzed by electrophoresis and can resolve contrasting alleles. Currently SSR are regarded as a marker of choice because of the high level of informativeness, co-dominance, wide spread in eukaryotic genomes, and ease of amplification by standard PCR technique.

Using these molecular markers, more than 960 QTL associated with various traits have already been mapped, but only a few of them have been validated. The reported QTL cannot be used in the molecular breeding programs until they are confirmed across populations and environments, because results may vary from cross to cross (Shoemaker et al., 1996). On the basis of consistent results from various crosses, a more complete and useful mapping of an existing locus can be made. Despite the importance of soy protein, no QTL have been reported for protein quality including amino acid composition. Therefore, a need exists to continue to map protein and other seed quality, and agronomic traits to particular positions on the chromosomes to facilitate future breeding programs to adopt MAS. Focusing on protein concentration and quality, I

was interested to discover genomic regions as QTL associated with nitrogen accumulation during reproductive stages, protein concentration, storage protein fractions, amino acid composition, seed size, oil content, and agronomic traits. This research was directed to achieve these objectives.

References

- Carlson, J.B., and N.R. Lersten. 2004. Reproductive morphology, p. 59-95, In H. R. Boerma and J. E. Specht, eds. Soybeans: Improvement, production and uses. Crop Science Society of America, Madison, WI, USA.
- Cregan, P.B., T. Jarvik, A.L. Bush, R.C. Shoemaker, K.G. Lark, A.L. Kahler, N. Kaya, T.T. VanToai, D.G. Lohnes, and J. Chung. 1999. An integrated genetic linkage map of the soybean genome. *Crop Sci* 39:1464-1490.
- Hymowitz, T., and R.J. Singh. 1987. Taxonomy and speciation, p. 23-48, In J. R. Wilcox, ed. Soybeans: Improvement, production and uses, Second ed. Crop Science Society of America, Madison, WI, USA.
- Lersten, N.R., and J.B. Carlson. 2004. Vegetative morphology, p. 15-57, In H. R. Boerma and J. E. Specht, eds. Soybeans: Improvement, production and uses. Crop Science Society of America, Madison, WI, USA.

- Liu, K. 1997. Soybeans: chemistry, technology and utilization Chapman and Hall.
- Mounts, T.L., W.J. Woulf, and W.H. Martinez. 1987. Processing and utilization, p. 819-866, In J. R. Wilcox, ed. Soybeans: Improvement, production and uses, Second ed. Crop Science Society of America, Madison, WI, USA.
- Palmer, R.G., and T.C. Kilen. 1987. Qualitative genetics and cytogenetics, p. 135-209, In J. R. Wilcox, ed. Soybeans: Improvement, production and uses, Second ed. Crop Science Society of America, Madison, WI, USA.
- Palmer, R.G., T. Hymowitz, and R.L. Nelson. 1996. Germplasm diversity within soybean, p. 1-35, In D. P. S. Verma and R. C. Shoemaker, eds. Soybean: Genetics, Molecular Biology and Biotechnology. CAB International, Wallingford, UK.
- Poehlman, J.H., and D.A. Sleper. 1995. Breeding field crops. Fourth ed. Iowa State University Press, Ames, IA, USA.
- Shoemaker, R.C., K.M. Polzin, L.L. Lorenzen, and J.E. Specht. 1996. Molecular genetic mapping of soybean, p. 37-56, In D. P. S. Verma and R. C. Shoemaker, eds. Soybean: Genetics, Molecular Biology and Biotechnology. CAB International, Wallingford, UK.
- Song, Q.J., L.F. Marek, R.C. Shoemaker, K.G. Lark, V.C. Concibido, X. Delannay, J.E. Specht, and P.B. Cregan. 2004. A new integrated

genetic linkage map of the soybean. Theor Appl Genet 109:122 -
128.

Soytech Inc. 2004. Oilseed statistics. Soytech Inc., Bar Harbor, ME.

PART 2

Genomic Regions Governing Soybean Seed Nitrogen Accumulation

This chapter is a lightly revised version of paper by the same name that has been published in ***Journal of American Oil Chemists' Society* 2004 81: 77-81** by D.R. Panthee, V.R. Pantalone, C.E. Sams, A.M. Saxton, D.R. West and W.E. Rayford.

My use of “our” in this chapter refers to my co-authors and myself. My primary contribution to this paper include (1) organization, development and management of all aspects of the project, (2) collecting all molecular data using polymorphic molecular markers, (3) collecting all the field data (4) analyzing for nitrogen and amino acids composition, (5) analyzing all data, and (6) all of the writing except for some minor revisions.

Abstract

Nitrogen accumulation for seed protein takes place in soybean [*Glycine max* (L.) Merr.] during the reproductive stages of development. The purpose of this study was to determine genotypic differences in seed nitrogen accumulation, and to identify genomic regions controlling nitrogen accumulation in soybean during R₅, R₆ and R₇ growth stages. A population of 105 F_{6:8} recombinant inbred lines (RIL) developed from a cross of N87-984-16 x TN93-99 was utilized. The RIL were grown at the University of Tennessee, Knoxville Experiment Station in a randomized complete block design with three replications in 2002. Pod samples were taken at R₅, R₆ and R₇ growth stages and seed nitrogen was determined.

A significant ($P < 0.05$) difference among genotypes was found for nitrogen accumulation at all three growth stages. Single factor ANOVA revealed that quantitative trait loci (QTL) governing nitrogen accumulation in soybean seed were distributed in the linkage groups A2, B2, D1a, D1b, E, G and M. Phenotypic variation explained by an individual QTL ranged from 5 to 11.6%. These QTL may provide useful marker assisted selection opportunities for soybean protein improvement.

KEY WORDS: nitrogen accumulation, protein accumulation, QTL, protein

Introduction

Soybean [*Glycine max* (L.) Merr.] seeds typically contain approximately 40% protein on a dry weight basis, making soybean one of the highest protein containing crop species. Nitrogen (N) is an important constituent of seed storage protein and is accumulated in soybean seed during reproductive stages of development. This process is essential for seed protein and seed yield formation. The rate of storage protein accumulation can be estimated from the rate of N accumulation.

Understanding the physiology of N accumulation and protein synthesis should allow more efficient breeding to develop high protein soybean lines. According to Shibles *et al.* (1994), Sources of N include uptake of soil nitrate during seed growth, fixation of atmospheric N_2 during seed growth; and redistribution of N accumulated in vegetative tissues before

seed growth begins. Staswick *et al.* (2001) determined that large quantities of N are also mobilized from mature organs to developing seeds. Imsande (2001) found that leaf proteins are hydrolyzed and the salvaged N is translocated to the developing pods and seeds during soybean pod filling. An increase in the protein content of the cotyledon commences as early as 15 days after flowering and continues through to maturity. The most rapid rate of accumulation of total protein in cotyledons is between 25 to 40 days after flowering (Gayler and Sykes, 1981). Accumulation of N into growing seeds and formation of seed storage proteins, which contribute to soybean seed yield has an important economic aspect.

Kumudini *et al.* (2002) demonstrated a genotypic effect in seed N accumulation in soybean, where an ancestral set of cultivars (Pagoda and Mandarin) was superior in this process compared to later cultivars (Maple Glen and OAC Bayfield). Recently, Grandgirard (2002) further proved that seed N accumulation rate in soybean is determined by the genotype. However, there are no reports available about the genomic regions or quantitative trait loci (QTL) controlling this process. Soybean storage protein is inherited as a quantitative trait (Nakasathien *et al.*, 2000), which is influenced by the growing environment. At the same time, there is a potential for interaction of genes that mediate metabolic pathways. Therefore, it is important to consider genotypic effects on seed

composition to provide stable production. For this reason, knowledge of genotypic differences in the timing and rates of constituent deposition (like N accumulation) during seed development is important (Wilson, 2004). Once the rate of N accumulation in the seed is understood in a given genotype, management practices can be adjusted to obtain the maximum N accumulation in soybean seed. QTL analyses can detect important genomic regions governing N accumulation. Furthermore, knowledge of specific QTL provides opportunities for marker-assisted selection in cultivar development programs. Therefore, to gain more insight, this study was conducted to determine if there are genotypic differences in soybean for seed N accumulation, and to identify genomic regions controlling this process in soybean.

Materials and Methods

Plant materials

A total of 105 F_{6:8} recombinant inbred lines (RIL) were developed from a cross of N87-984-16 x TN93-99. The N87-984-16 parent is one of two sister lines, whose blend constitutes the high protein commercial variety 'Prolina' (Burton et al., 1999). The TN93-99 parent is a high yielding and well adapted University of Tennessee breeding line and is currently registered as a germplasm (Pantalone et al., 2003). All RIL, the two parents, and the check cultivars 'Hutcheson', '5002T' and '5601T' were planted in a randomized complete block design with three replications at

the Knoxville Experiment Station of the University of Tennessee, Knoxville, TN in May 2002. Each line was planted in a four-row plot of 6 m length with a spacing of 75 cm between rows.

Sample collection

Pods were collected at R₅, R₆ and R₇ growth stages (Fehr and Caviness, 1977). Fehr and Caviness (1977) regarded the R₂ stage as 'days to flowering'. Once days to flowering for all the lines were recorded, days to R₅, R₆ and R₇ growth stages were estimated for a tentative visit for pod collection. Pod collection was estimated to occur at five, seven and nine weeks after R₂ for R₅, R₆ and R₇ growth stages respectively. Actual sample collection was done when the individual genotype reached the reproductive growth stages as specified in Fehr and Caviness (1977). At R₅, R₆ and R₇, pods were sampled from the 7th to 9th node of the plant to maintain uniformity. Approximately 25 pods were collected at R₅, 15 pods at R₆ and 10 pods at R₇ so that at least 1 g of dried ground sample would be available. The seeds were separated from the pods and dried at 60°C for 24 hours. Dried samples were ground with a coffee grinder for about 20 s and transferred to 2 mL Eppendorf tubes.

Nitrogen determination

Nitrogen content of the sample was determined using a LECO CHN 2000 Analyzer (LECO Corporation, Lakeview Ave., MI) following AOAC Official

Method 990.03 (AOAC, 1995). The principle involved in this analysis is that nitrogen freed by combustion at high temperature in pure oxygen is measured by thermal conductivity detection and can be converted to equivalent protein by an appropriate numerical factor (6.25 in the case of soybean protein).

Amino acid analysis

Ten grams of soybean seeds were ground in a water-cooled (20°C) Knifetec 1095 Sample Mill (FOSS Tecator, S-26321, Hogana, Sweden) for 20 seconds. This process produced soybean flour with a uniform particle size. The ground sample was analyzed via near-infrared reflectance, using a FOSS 6500 instrument and amino acid prediction equation (FOSS North America, Eden Prairie, MN).

DNA extraction and polymerase chain reaction

DNA was extracted from all RIL and the two parental lines by following Qiagen Plant Easy DNA Extraction Kit (Qiagen Hilden, Germany). Polymerase chain reaction (PCR) consisted of 7.4 µL of ddH₂O, 1 µL of 10X PCR Buffer, 1 µL of 2 mM dNTPs mixture (Pharmacia, Piscataway, NJ), 0.5 µL of 20 µM forward and reverse primer, 0.1 µL of 5 units/µL KlenTaq (Ab Peptides Inc., St. Louis, MO) and 2 µL of 20 ng/µL template DNA. The PCR was performed in a 96-well MBS Hybaid thermocycler (Hybaid, Franklin, MA). PCR conditions were a) 94°C for five minutes

followed by 35 cycles at 94°C for denaturation for 25 s, b) 47°C for annealing for 30 s, c) 72°C for 25 s for extension, and one last cycle at 72°C for final extension for five minutes. Parents were screened with a total of 568 (ATT)_n type of simple sequence repeat (SSR) genetic markers, which spanned all 20 molecular linkage groups (LG) of soybean (Cregan et al., 1999). The DNA sequence information for soybean SSR markers is publicly available from the website (www.soybase.org) (verified December 16, 2004). A total of 86 SSR markers, distributed among all 20 LG were found to be polymorphic between the parents and among the RIL, and were used in QTL analyses.

A 6% non-denaturing polyacrylamide gel electrophoresis (PAGE) was used to separate the PCR product. The PAGE consisted of 28.5 mL (40% bis-acrylamide solution), 160.2 mL (0.60X TBE buffer), 1.33 mL (10% APS) and 66.5 µL (TEMED). Two µL loading buffer was added to PCR product and 10 µL sample was loaded in the gel. Running buffer was 0.5X TBE. The gel was run at a constant 300 V for 3 h. A fan was used to keep the glass plates cool during gel running. The bands were visualized by staining the gel with 50 µL ethidium bromide (10 mg/mL) and photographing it under exposure to UV light, and scored using 1 for parent 1 (N87-984-16), 2 for heterozygote and 3 for parent 2 (TN93-99).

Data analysis

Nitrogen content data were analyzed using SAS software (SAS Institute Inc, 2002). Associations between molecular markers and nitrogen accumulation at various growth stages (R_5 , R_6 and R_7), and amino acids were analyzed with SAS using single factor ANOVA. QTL position was determined by analyzing the data with MAPMAKER/EXP 3.0 (Lincoln et al., 1993) and QTL Cartographer 2.0 (Wang et al., 2003).

Results and Discussion

Recombinant inbred lines are fixed in the homozygous state at nearly all loci. Thus, differences in a measured trait (such as seed N content), which are associated with allelic differences at specific DNA marker loci, reveal the underlying genomic regions governing that trait. Seed N content varied from 4.63 to 9.91% (Table 2.1) among the soybean RIL determined at different growth stages. The frequency distribution of all the RIL is presented in Figure 2.1. The RIL differed significantly ($P < 0.05$) for N accumulation indicating that some had better genetic potential than others for this character suggesting an opportunity to select specific RIL for high nitrogen accumulation. This finding is consistent with the results of Grandgirard (2002).

Table 2.1: Descriptive statistics for nitrogen and protein content on a dry weight basis in a soybean RIL population (n=105) measured at three reproductive growth stages.

Descriptors	R ₅ Stage		R ₆ Stage		R ₇ Stage	
	Nitrogen	Protein	Nitrogen	Protein	Nitrogen	Protein
Minimum	5.02	31.38	4.63	28.91	5.11	31.92
Mean	6.29	39.29	6.34	39.60	6.67	41.71
Maximum	8.25	51.53	9.91	61.94	7.79	48.66
Standard deviation	0.38	2.39	0.52	3.25	0.36	2.24

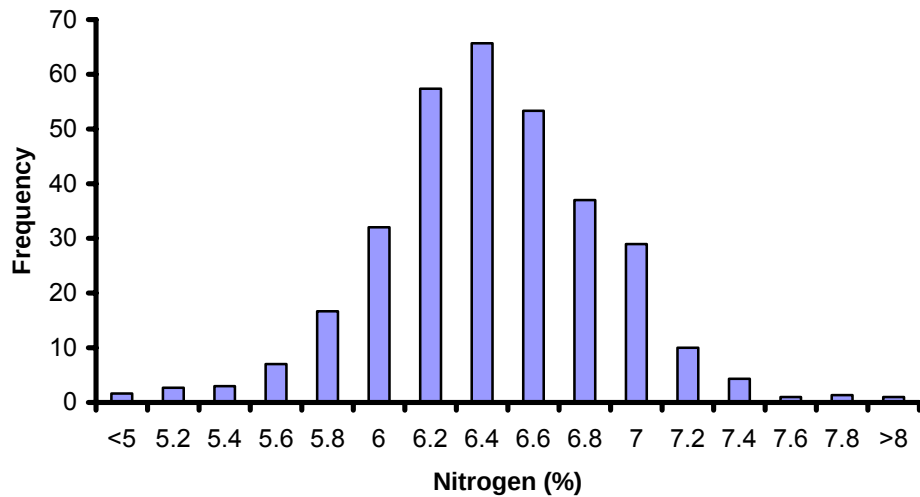


Figure 2.1: Frequency distribution for a population of soybean recombinant inbred lines for seed nitrogen content class means (constructed at intervals of 0.2% nitrogen), averaged over three reproductive growth stages.

In our study, there was a gradual increase in seed N content at different growth stages. The seed N content differed significantly ($P < 0.05$) between R_5 , R_6 and R_7 growth stages. Mean N contents were 6.28% at R_5 , 6.35% at R_6 and 6.68% at R_7 . Most of the N accumulation took place within five weeks after flowering in soybean. This has management implications for efforts targeting increases in protein or specific amino acid content via nutrient amendments (Nakasathien et al., 2000; Paek et al., 2000; Sexton et al., 1998). Depending upon the time required for N uptake by the growing soybean plant, time of N application can be adjusted. Our observations verify the findings of Gayler and Sykes (1981) that the majority of seed N accumulates by R_5 stage. They found that the protein fraction starts accumulating 15 to 17 days after flowering (DAF) and reaches a maximum by 25 to 30 DAF. They also found that the majority of protein accumulates within five weeks. Grandgirard (2002) mentioned that seed N accumulation rate is highly correlated with the instantaneous available N concentration. Lhuillier-Soundele *et al.* (1999) also found that seed N accumulation was positively related to available N in vegetative parts in pea (*Pisum sativum* L.).

Cregan et al. (1999) developed a large set of SSR molecular markers for soybean, which are currently well mapped to 20 molecular linkage groups, which correspond to the 20 pairs of homologous chromosomes in soybean. After screening a total of 568 SSR markers, from which 86

markers were identified as exhibiting polymorphic DNA banding patterns among the RIL, we identified seven markers linked to QTL governing N accumulation at the R₅ stage in soybean. Those QTL were distributed among five linkage groups (A2, B2, D1b, E and G). The variation among RIL for N accumulation explained by the QTL ranged from 5.9 to 10.7% (Table 2.2). Genomic region near Satt274 (LG D1b) was identified to have a major QTL for seed N content at the R₅ growth stage.

There were three markers linked to QTL governing N accumulation in seeds at the R₆ growth stage. At R₆, in addition to linkage group A2 and D1b, markers from linkage group M were also detected. The strongest QTL (linked to Satt459 on LG D1b) explained 9.3% of the variation among RIL for N accumulation. At R₇, there were seven markers associated with QTL from four linkage groups (B2, D1a, E and G) controlling N accumulation. The Satt263 marker was linked with the strongest QTL located on LG E and explained 11.6% of variation among RIL for N accumulation. There are few reports available on genomic regions controlling seed nitrogen content in crop species. Bezant *et al.* (1997) identified eight QTL controlling seed N content in barley using RFLP markers. Our results identifying QTL for seed N accumulation at various growth stages in soybean have important implications in soybean crop management and cultivar development.

Table 2.2: Molecular markers and molecular linkage groups for QTL governing nitrogen accumulation in developing seeds of soybean at three reproductive growth stages.

Markers	Linkage group†	Probability	R ² (%)	Allele mean		Difference (P1-P2)	Genetic Gain
				N (%)			
				N87 (P1)	TN93 (P2)		
R ₅ growth stage							
Satt177	A2	0.027	6.1	6.27	6.38	-0.11	1.75
Satt133	A2	0.027	6.6	6.24	6.35	-0.11	1.76
Satt556	B2	0.033	5.9	6.38	6.26	+0.12	1.92
Satt274	D1b	0.004	10.7	6.24	6.39	-0.15	2.40
Satt185	E	0.011	9.0	6.34	6.22	+0.12	1.93
Satt263	E	0.026	7.0	6.35	6.23	+0.12	1.93
Satt235	G	0.019	7.1	6.37	6.25	+0.12	1.92
R ₆ growth stage							
Satt437	A2	0.015	8.2	6.27	6.42	-0.15	2.39
Satt459	D1b	0.100	9.3	6.44	6.28	+0.16	2.55
Satt201	M	0.028	6.5	6.33	6.46	-0.13	2.05
R ₇ growth stage							
Satt168	B2	0.049	5.0	6.75	6.64	+0.11	1.66
Satt383	D1a	0.025	5.9	6.75	6.64	+0.11	1.66
Satt268	E	0.019	7.0	6.63	6.75	-0.12	1.81
Satt263	E	0.004	11.6	6.62	6.77	-0.15	2.27
Satt45	E	0.004	9.9	6.62	6.76	-0.14	2.11
Satt570	G	0.028	6.3	6.74	6.62	+0.12	1.81
Satt131	G	0.023	6.3	6.77	6.65	+0.12	1.80

†Linkage group reported by Cregan *et al.* (1999).

It should be noted that the present study was conducted only in one environment, but with sampling over three replication and three growth stages. In order to confirm QTL and to identify environmentally stable QTL (Qiu et al., 1999), additional molecular mapping can be performed in multiple locations and in additional populations by other researchers. Although the results presented in this study are limited to one environment, it should be noted that soybean protein is considered to be a highly heritable trait; i.e. genetic factors (rather than environmental factors) primarily govern measurable variation for protein. Thus, the present study has provided potentially important information for marker-assisted selection to increase protein concentration in soybean by identifying QTL associated with N accumulation at reproductive growth stages.

It is interesting to note that there were few common linkage groups associated with N accumulation across different growth stages (Table 2.2). This suggests that the same gene may not express throughout reproductive development. Rather, total seed N accumulated may be the result of cumulative effects of several QTL. There were significant differences among RILs for amino acid composition (data not shown), which enabled detection of QTL for specific amino acids. This was likely because the N87-984-16 parent (Prolina) has greater levels of the sulfur-containing amino acids (cysteine and methionine) than normal cultivars (Wilson, 2004). The QTL for seed N accumulation found at different

growth stages were found to also govern specific amino acids in this population (Table 2.3). For example, QTL for methionine were detected near Satt133 (LG A2) and Satt 235 (LG G), which were also R₅ seed N QTL; near Satt459 (LG D1b) and Satt201 (LG M), which were also R₆ seed N QTL; and near Satt45 (LG E) and Satt268 (LG E), which were also R₇ seed N QTL (Tables 2.2 and 2.3). This suggests that different amino acids are synthesized throughout different growth stages, or that assembly of specific amino acids requires different genes, which are regulated during reproductive growth stages.

Table 2.3: Soybean seed nitrogen accumulation QTL at three reproductive growth stages, which were also, detected as amino acid QTL.

Growth stage	Linkage group†	Marker	Amino acid
R ₅	A2	Satt177	Cys
	A2	Satt133	Met, Gly, Pro, Val, Iso, Phe, His, Tyr
	D1b	Satt274	Cys, Thr
	E	Satt185	Asp, Ser, Arg, His
	E	Satt263	Asp, Glu, Ser, Val, Iso, Leu, Phe, His, Tyr
	G	Satt235	Met, Arg
R ₆	A2	Satt437	Gly, Val, Iso
	D1b	Satt459	Met, Glu, Gly, Iso
	M	Satt201	Met, Tyr
R ₇	B2	Satt168	Thr
	E	Satt268	Met, Arg
	E	Satt263	Asp, Glu, Ser, Val, Arg, Iso, Leu, Phe,
	E	Satt45	His, Tyr
	G	Satt570	Met, Asp, Glu, Arg
	G	Satt131	Thr, Arg, Pro

†Linkage group reported by Cregan *et al.* (1999).

All the QTL detected governing seed N accumulation were located on seven linkage groups (A2, B2, D1a, D1b, E, G and M). Brummer *et al.* (1997) mapped QTL controlling protein content in eight different soybean populations using RFLP markers. They identified QTL for protein linked to markers located on nine different linkage groups including A2, B2, D1, E and G, which are consistent with our observations. Different QTL may be identified in different populations because of genetic background effects. Qui *et al.* (1999) used a population from Peking x Essex for mapping various characters including protein. They used RFLP markers and found QTL associated with protein on LG F and H. SSR markers, associated with protein QTL were found to be located on linkage groups D1a and M by Csanadi *et al.* (2001). This is consistent with our findings.

Our study has established that N accumulation is affected by the genotype of soybean and that there are genomic regions controlling this process. The QTL identified in this study provide potentially useful marker-assisted selection opportunities for soybean protein improvement.

References

- AOAC. 1995. Protein (crude) in Animal Feed Method 990.03, Arlington, VA.
- Bezant, J.H., D.A. Laurie, N. Pratchett, J. Chojecki, and M.J. Kearsey. 1997. Mapping of QTL controlling NIR predicted hot water extract

- and grain nitrogen content in a spring barley cross using marker-regression. *Plant Breeding* 116:141-145.
- 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., T.E. Carter, and R.F. Wilson. 1999. Registration of Prolina soybean. *Crop Sci* 39:294-295.
- Cregan, P.B., T. Jarvik, A.L. Bush, R.C. Shoemaker, K.G. Lark, A.L. Kahler, N. Kaya, T.T. VanToai, D.G. Lohnes, and J. Chung. 1999. An integrated genetic linkage map of the soybean genome. *Crop Sci* 39:1464-1490.
- Csanadi, G., J. Vollmann, G. Stift, and T. Lelley. 2001. Seed quality QTLs identified in a molecular map of early maturing soybean. *Theor Appl Genet* 103:912-919.
- Fehr, W.R., and C.E. Caviness. 1977. Stages of soybean development. Iowa State University, Ames, IA.
- Gayler, K.R., and G.E. Sykes. 1981. β -conglycinins in developing soybean seeds. *Plant Physiol* 67:958-961.
- Grandgirard, D. 2002. Analysis and modeling of soybean (*Glycine max* L.) seed nitrogen concentration: Relation between nitrogen remobilization to the seeds and crop yield and quality. Ph.D. Thesis, Institut National de la Recherche Agronomique (INRA), Dijon, France.

- Imbande, J. 2001. Selection of soybean mutants with increased concentrations of seed methionine and cysteine. *Crop Sci* 41:510-515.
- Kumudini, S., D.J. Hume, and G. Chu. 2002. Genetic improvement in short-season soybeans. II. Nitrogen accumulation, remobilization, and partitioning. *Crop Sci* 42:141-145.
- Lhuillier-Soundele, A., N.G. N.G. Munier-Jolain, and Ney. 1999. Influence of nitrogen availability on seed nitrogen accumulation in pea. *Crop Sci* 39:1741-1748.
- Lincoln, S.E., M.J. Daly and E.S. Lander. 1993. Constructing genetic linkage maps with MAPMAKER Version 3.: A tutorial and reference manual. Release 3rd edition. Whitehead Institute for Biomedical Research, Cambridge, MA.
- Nakasathien, S., D.W. Israel, R.F. Wilson, and P. Kwanyuen. 2000. Regulation of seed protein concentration in soybean by supra-optimal nitrogen supply. *Crop Sci* 40:1277-1284.
- Paek, N.C., P.J. Sexton, S.L. Laeve, and R. Shibles. 2000. Differential accumulation of soybean seed storage protein subunits in response to sulfur and nitrogen nutritional sources. *Plant prod sci* 3:268-274.
- Pantalone, V.R., F.L. Allen, and D. Landau-Ellis. 2003. Registration of 'TN93-99' Soybean germplasm. *Crop Sci* 43:1137.

- Qiu, B.X., P.R. Arelli, and D.A. Sleper. 1999. RFLP markers associated with soybean cyst nematode resistance and seed composition in a 'Peking' x 'Essex' population. *Theor Appl Genet* 98:356-364.
- SAS Institute Inc. 2002. SAS guide for personal computers. 9th ed. SAS Institute, Cary, NC.
- Sexton, P.J., S.L. Naeve, N.C. Paek, and R. Shibles. 1998. Sulfur availability, cotyledon nitrogen:sulfur ratio, and relative abundance of seed storage proteins of soybean. *Crop Sci* 38:983-986.
- Shibles, R.M., D.N. Sundberg, and S.L. Naeve. 1994. Physiological constraints limiting seed protein in normal and high seed protein soybean genotypes. *Agron Abstract* 1994:151.
- Staswick, P.E., Z. Zhang, T.E. Clemente, and J.E. Specht. 2001. Efficient down-regulation of the major vegetative storage protein genes in transgenic soybean does not compromise plant productivity. *Plant Physiol* 127:1819-1826.
- Wang, S, C, .J. Basten and Z.B. Zeng. 2003. Windows QTL Cartographer. Release V2.0. North Carolina State University, Raleigh, NC.
- Wilson, R.F. 2004. Seed composition, *In* H. R. Boerma and J. E. Specht, eds. *Soybeans: Improvement, Production and Uses*, 3rd Edition ed. ASA, CSSA, SSA, Madison, WI.

PART 3

Quantitative Trait Loci for Seed Protein and Oil Concentration, and Seed Size in Soybean

This chapter is a lightly revised version of paper by the same name that has been submitted to the ***Crop Science*** in 2004 by D.R. Panthee, V.R. Pantalone, D.R. West, A.M. Saxton and C.E. Sams

My use of “our” in this chapter refers to my co-authors and myself. My primary contribution to this paper include (1) organization, development and management of all aspects of the project, (2) collecting all molecular data using polymorphic molecular markers, (3) analyzing seed samples for protein, oil and seed size, (4) analyzing all data, and (5) all of the writing except for some minor revisions.

Abstract

Soybean [*Glycine max* (L.) Merr.] is an important crop because of its high oil and protein concentration. However, there is an inverse relationship between seed protein and oil concentration making it difficult to improve both traits simultaneously by conventional breeding methods. Molecular breeding may be helpful to overcome such a barrier if specific genomic regions controlling these two traits could be identified to facilitate a balanced accumulation of desirable alleles. The objective of this study was to determine genotypic differences in soybean for protein and oil concentration, and seed size and to identify genomic regions governing these traits. To achieve this objective, 101 F₆-derived recombinant inbred lines (RIL) from a population developed from a cross of N87-984-16 x

TN93-99 were used. Protein and oil were determined using a near infrared transmittance spectroscopy (NITS) instrument on whole seed harvested from a total of six environments over two years of testing. A significant ($P < 0.05$) difference among the RIL was found for all three seed traits. Heritability estimates on an entry mean basis for protein and oil concentration, and seed size were 0.66, 0.54 and 0.71, respectively. A total of 94 polymorphic simple sequence repeat (SSR) molecular genetic markers were screened from DNA of the RIL. Single factor ANOVA was used to identify candidate QTL, which were then confirmed by composite interval mapping via QTL Cartographer. Two QTL (near Satt274 and Satt570) for protein concentration, seven QTL (near Satt274, Satt293, Satt317, Satt420, Satt459, Satt562 and Satt570) for oil concentration and eight QTL (near Satt002, Satt147, Satt184, Satt185, Satt255, Satt268, Satt459 and Satt540) for seed size were identified. The individual QTL explained 12.8-24.9%, 9.4-18.1% and 10.0-39.9% of the phenotypic variation for protein and oil concentration, and seed size, respectively. Thus, we identified major loci governing each of these seed traits. Three improved protein lines were developed from this population and released as germplasms by the Tennessee Agricultural Experiment Station: TN03-350, TN04-5321 and TN04-5363. These germplasms and QTL identified in this study may be useful for improving soybean seed quality.

Introduction

Soybean [*Glycine max* (L.) Merr.] typically contains 400 g kg⁻¹ protein and 200 g kg⁻¹ oil on a seed dry weight basis (Wilson, 2004). Soybean is grown as an oil and protein source for human and livestock nutrition, and for use in pharmaceuticals, and industrial products in North and South America, Europe, and Asia (Liu, 1997). Food items prepared from soybean include textured soy protein, soy isolates, soymilk, tofu, natto, miso, tempeh, soy sprouts and soy sauce. Seed traits such as protein concentration and seed size have important roles in determining the quality of these soy food items. Considering the nutritional and economic importance of protein and oil in soybean, attempts have been made to increase their concentrations in soybean seed. However, there is typically a negative correlation between protein and oil concentration in soybean seed (Burton, 1987). It is, however, possible to increase the concentration of one at the expense of the other.

Hardiness, brittleness and gumminess are the physical properties of tofu, which determine the overall quality. Protein concentration, particularly glycinin, determines these properties in tofu whereas seed size determines the quality of soy sprouts (Liu, 1997). Generally small seeded soybeans are desired for high quality soy sprouts and natto production, and combining higher protein with large seed size is desirable for tofu production. Genomic regions for seed size could be used in marker-

assisted selection (MAS) along with protein and oil QTL for soy food applications.

Despite moderately high heritabilities, it is difficult to improve seed traits simultaneously, particularly protein and oil concentration. The MAS approach might be useful towards achieving specific goals if genomic regions controlling protein and oil concentration, and seed size could be identified more precisely to improve selection indices.

Several researchers have paved the way towards this goal. Recently, SSR molecular markers associated with seed size in soybean were identified (Hoeck et al., 2003). They used three populations to identify QTL associated with seed size. They found 12 QTL in population 1 (A97-775019 x A96-492041), 16 in population 2 (A97-775006 x S12-49) and 22 in population 3 (A97-775026 x A96-492058). However, they found only 13 unique QTL associated with seed size. Orf et al. (1999) also used three populations (Minsoy x Noir 1, Archer x Minsoy, and Noir 1 x Archer) of soybean to study QTL associated with seed size. They found seven QTL in each of two populations, and two QTL in one population. Mian et al. (1996) identified 16 QTL associated with seed size in two populations of soybean, of which 12 QTL were unique. Maughan et al. (1996) developed an inter-specific population of soybean and identified three unique QTL

associated with seed size. Breeders use of such QTL is more likely after they are confirmed in diverse populations grown in diverse environments.

Qiu et al. (1999) found two RFLP markers associated with QTL controlling protein, and only one marker controlling oil concentration in a population derived from 'Peking' x 'Essex'. Since the total phenotypic variation explained by the QTL was around 30%, they assumed that there should be additional QTL determining these traits, which could not be detected because of background genetic effects of the population in their experiment. Csanadi et al. (2001) mapped QTL for protein, oil and seed size in an early maturing soybean population developed from the cross of cultivars 'Ma Belle' x 'Proto'. They used SSR, RAPD and RFLP to map the QTL and found four QTL for protein, three for oil and eight for seed size. Mansur et al. (1996) mapped QTL for seed size, protein and oil concentration in a population developed from 'Minsoy' x 'Noir 1'. They found three markers associated with seed size, protein and oil content each.

Currently, there are at least 53 QTL associated with oil content, 61 with protein concentration and 66 with seed size (Hyten et al., 2004). However, most of the QTL are not confirmed (Pantalone et al., 2004). Furthermore, the populations have been derived using common parents or close parents in many of the cases (Hyten et al., 2004). However, it is important

to identify QTL in key pedigree lines and in as many populations as feasible derived from modern parents and grown in multiple environments to verify stable QTL. According to Brummer et al. (1997), QTL identified only in one environment and one population may not be stable in others because of interactions with the environments or because of genetic background effects. The objective of the present study was to detect QTL associated with seed protein and oil concentration, and seed size in a soybean population targeted for protein enhancement.

Materials and Methods

Plant materials

A total of 101 F_6 -derived recombinant inbred lines (RIL) were developed from a cross of N87-984-16 x TN93-99. The N87-984-16 parent is one of the two F_8 -derived sister lines of N87-984, a blend of which constitutes the high protein commercial cultivar 'Prolina' (Burton et al., 1999). Prolina was selected from the first cycle of a recurrent selection population designated NRS5. The population was originated from the matings of 10 high protein lines with the cultivars 'Bragg', 'Ransom' and 'Davis' (Burton et al., 1999). N87-984-16 also has large seed size. TN93-99 is an F_4 -derived, high yielding, above average oil concentration breeding line and is currently registered as germplasm GP-280. This line was developed from a triparental crossing (TN85-55 x TN5-85) x Hutcheson (Pantalone et al., 2003). In our study, crossing was made during summer of 1998. F_1 seeds

were harvested in October 1998 and F_1 single plants were grown in Costa Rica during winter 1998-99. Generations were advanced until the F_5 stage via the single seed descent method (Brim, 1966) in Costa Rica and $F_{5:6}$ seeds were obtained in May 2000. Approximately 300 $F_{5:6}$ single plants were grown at Knoxville, TN and 101 random plants of similar Tennessee adapted maturity were sampled as a source population for this study. All 101 lines and parents were planted in 6 m length two-row plots with three replications in a randomized complete block design (RCBD) on sequatchie silt loam soil at Knoxville Plant Science Farm (KPSF) of the Knoxville Experiment Station (Knoxville, TN) in a preliminary trial in 2001. All RIL, the two parents, and three checks ('Hutcheson', '5002T' and '5601T') were planted in a RCBD with three replications at three locations: on sequatchie silt loam soil at the Knoxville and sequatchie silt loam soil at Holston unit of the Knoxville Experiment Station of the University of Tennessee, and on sequatchie silt loam soil at the Ames Plantation near Grand Junction, TN in 2002 and 2003. Each line was planted in a four row plots of 6 m length with spacing of 75 cm between rows.

Phenotypic traits

Approximately 25 g of seed sample was analyzed as whole beans for protein and oil concentration per plot using an Infratec 1225 Food and Feed Grain Analyzer (Tecator, North America) at USDA-NCAUR, Peoria, IL. The Infratec analyzer is a near infrared transmittance instrument and is

used for simultaneous determination of feed constituents. The measurements were based on absorbance of electromagnetic radiation in the near infrared region of the spectrum. Seed size was determined by counting and weighing 100 seeds per plot.

DNA extraction and polymerase chain reaction

DNA was extracted from the RIL and parental lines utilizing Qiagen Plant Easy DNA Extraction Kit (Qiagen Ilden, Germany). Polymerase chain reaction (PCR) consisted of 7.4 μL of ddH₂O, 1 μL of 10X PCR Buffer, 1 μL of 2 mM dNTPs mixture (Pharmacia, Piscataway, NJ), 0.5 μL of 20 μM forward and reverse primer, 0.1 μL of 5 units/ μL Klentaq (Ab Peptides Inc., St. Louis, MO) and 2 μL of 20 ng/ μL template DNA. The PCR was performed in a 96-well MBS Hybaid thermocycler (Hybaid, Franklin, MA). PCR conditions were a) 94°C for five minutes followed by 35 cycles at 94°C for denaturation for 25 s, b) 47°C for annealing for 30 s, c) 72°C for 25 s for extension, and one last cycle at 72°C for final extension for five minutes. Parents were screened for a total of 585 (ATT)_n type of simple sequence repeat (SSR) genetic markers (Cregan et al., 1999). The sequence information for the markers is publicly available from the Website (www.soybase.org), which was verified on December 16, 2004. A total of 94 polymorphic SSR markers were used in QTL analysis.

DNA gel electrophoresis

A 6% non-denaturing polyacrylamide gel electrophoresis (PAGE) consisted of 6% bis-acrylamide, 0.5% TBE buffer, 0.07% APS and 0.035% TEMED, which was measured as 28.5 mL of 40% bis-acrylamide, 160.2 mL of 0.593X TBE buffer, 1.33 mL of 10% APS and 66.5 μ L of TEMED was used to separate the PCR product. Two μ L of loading buffer (6X) were added to PCR products and a 10 μ L sample was loaded on the gel. The running buffer was 0.5X TBE. The gel was run at constant 300 V for 3 h. A fan was used to circulate air to keep the glass plates cool during gel running. Ethidium bromide (50 μ L, 10 mg mL⁻¹) was added to the running buffer in order to visualize the bands under exposure to UV light. Bands were scored using 1 to represent P₁ (N87-984-16), 2 to represent heterozygous, and 3 to represent P₂ (TN93-99) alleles for each primer locus.

Data analysis

Phenotypic data on protein, oil and seed size were analyzed using the PROC MIXED procedure of SAS software to determine the genotypic differences among the RIL (SAS Institute Inc, 2002). Location and replications were considered as random blocking factors in the model. Heritability of each trait in the population was estimated by the entry mean basis method (Nyquist, 1991) as follows:

$$h^2 = \frac{\sigma_g^2}{\sigma_g^2 + \left(\frac{\sigma_{ge}^2}{e}\right) + \left(\frac{\sigma^2}{re}\right)}$$

where, h^2 represents the heritability, σ_g^2 is genetic variance, σ_{ge}^2 is genotype x environment variance, σ^2 is error variance, r is number of replications and e is the number of environments. The MIXED procedure of SAS with COVTEST option was used for REML estimation of variance components for heritability calculations.

Phenotypic correlations were determined using the PROC CORR procedure and genetic correlations were determined using the following formula (Falconer and Mackay, 1996; Kearsy and Pooni, 1996):

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

where, r_G represents genetic correlation, Cov represents covariance, x represents the first trait, y represents the second trait and σ^2 is genetic variance. Cross products were generated using the MANOVA option of the GLM procedure.

Candidate QTL were identified by determining associations between molecular data and least square means for protein, oil and seed size were detected by analyzing the data using the GLM procedure (SAS Institute Inc, 2002) for each environment and combined over six environments. All

heterozygotes were excluded from the analysis because the RIL population was F_6 -derived, and our interest was to detect heritable alleles from pure lines. Marker order and distance were determined using Mapmaker/Exp 3.0 (Lincoln et al., 1993). Information from Mapmaker was used in QTL Cartographer (Wang et al., 2003) to verify the candidate QTL by composite interval mapping (CIM). We used the standard model *Zmapqtl* 6 in the CIM procedure with a 10 cM window size and a 2 cM walking speed. We performed 1000 permutations on all traits in each environment and in the combined data from six environments to establish the empirical LOD thresholds at the 5% probability level (Churchill and Doerge, 1994). Any QTL with a LOD score greater than the empirical LOD threshold level were accepted significantly associated with the trait.

Results and Discussion

Protein and oil concentration, and seed size

Protein and oil concentration, and seed size are three important traits in soybean determining seed quality. There was a significant difference ($p < 0.05$) among the RIL for protein and oil concentration, and for seed size in this population. Protein concentration ranged from 400 to 435 g kg⁻¹ with a mean of 416 g kg⁻¹ (Table 3.1). There was transgressive segregation for protein concentration with a few lines significantly greater than the high parent (P1) for this trait (Figure 3.1). The population was normally distributed with a standard deviation of 0.7 g kg⁻¹. Typical protein

Table 3.1: Descriptive statistics for seed protein and oil concentration, and for seed size in an F6-derived soybean population of N87-984-16 x TN93-99 averaged over six environments (2002 and 2003).

Trait	Minimum	Maximum	Mean	Std. Dev.	P1	P2	h^2 †
Protein (g kg ⁻¹)	400	435	416	0.7	431	401	0.66
Oil (g kg ⁻¹)	183	204	195	0.4	190	203	0.54
Seed size (mg seed ⁻¹)	125	165	144	0.9	150	141	0.71

P1 = N87-984-16; P2 = TN93-99; † Heritability on an entry mean basis

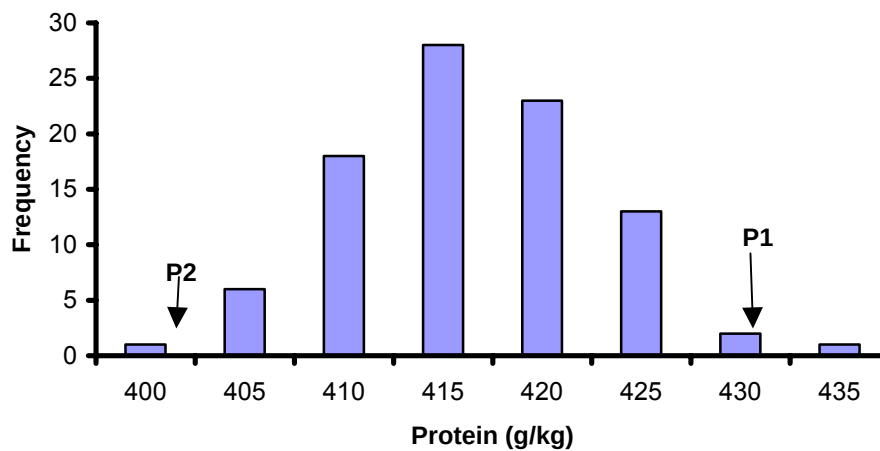


Figure 3.1: Frequency distribution of seed protein concentration (g kg⁻¹) in an F6-derived soybean population of N87-984-16 x TN93-99 averaged over six environments (2002 and 2003).

concentration in soybean is 400 g kg⁻¹; the population mean in this population was slightly higher than typical. Oil concentration ranged from 183 to 204 g kg⁻¹ with a population mean of 195 g kg⁻¹. Interestingly, there were no lines significantly greater than the high parent (P2) for oil concentration in this population (Table 3.1 and Figure 3.2). Seed size ranged from 125 to 165 mg seed⁻¹ with a population mean of 144 mg seed⁻¹. There was transgressive segregation for seed size in this

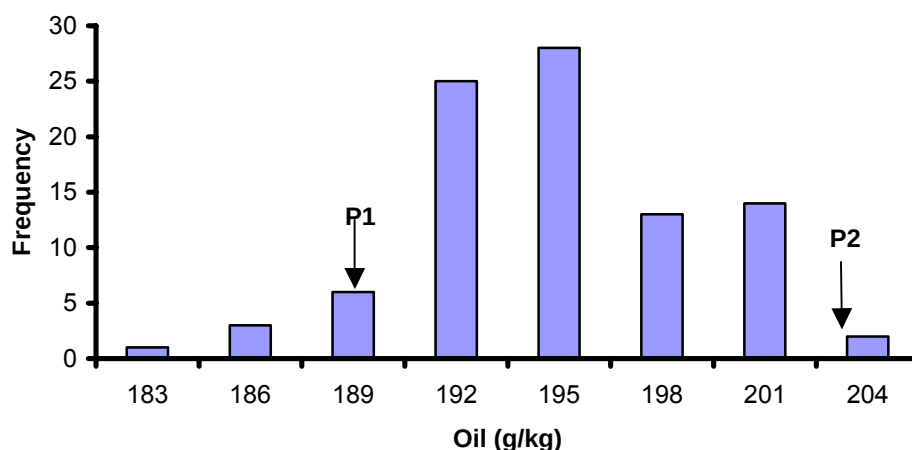


Figure 3.2: Frequency distribution of seed oil concentration (g kg⁻¹) in an F₆-derived soybean population of N87-984-16 x TN93-99 averaged over six environments (2002 and 2003).

population (Figure 3.3). Smaller seed size is preferred in Asian markets for quality sprouts and natto large seed size with high protein concentration is preferred for tofu production (Liu, 1997). Our population enables selection of lines for smaller or large seed size, however none would be the preferred sizes for either tofu or natto. There is a possibility of increasing the frequency of desirable alleles for protein or oil and for smaller or larger seed size from this population since the heritabilities for these traits are moderately high (Table 3.1).

The heritability for protein, oil and seed size found in this population indicated that most of the variation was genetic. Chung et al. (2003) have reported the heritability for protein and oil at 0.89 and 0.84 respectively. A

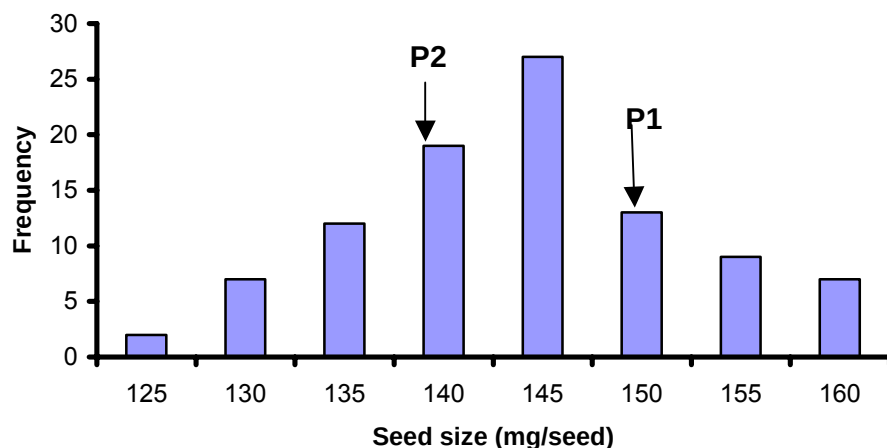


Figure 3.3: Frequency distribution of seed size (mg seed⁻¹) in an F6-derived soybean population of N87-984-16 x TN93-99 averaged over six environments (2002 and 2003).

range of heritability for protein and oil was reported as 0.56-0.92 and 0.07-0.81 respectively, depending upon the population (Brummer et al., 1997). The heritability observed in our population indicated that selection response would be reasonable for achieving genetic gain.

A wide range of seed weight (113-237 mg seed⁻¹) was reported in a study from the population aiming to produce soybean sprouts (Lee et al., 2001). Much less seed weight (as low as 81 mg seed⁻¹) was found in a different study (Hoeck et al., 2003), which used a small-seeded line as a parent. Our population was neither aiming to produce sprouts nor used a small seeded parent. Nevertheless the low seed weight in our study was close to that reported by Lee et al. (2001). They reported the heritability of seed weight as 0.95 and 0.83 respectively whereas in our population, it was

0.71. These studies indicated that seed size is a moderately high to highly heritable trait in soybean.

Large seed size with high protein concentration is desirable for tofu production (Alt et al., 2002). Larger seeded lines with high protein concentration observed in this population indicated that there is a possibility of developing tofu type lines utilizing this germplasm in further crossing programs with large seeded parent. Furthermore, positive genetic and phenotypic correlations were found between seed size and protein concentration in this population (Table 3.2). There was a negative phenotypic correlation between protein and oil content ($r = -0.59$, $P < 0.001$) but a positive phenotypic correlation between protein and seed size ($r = 0.50$, $P < 0.001$) in this population. There was close agreement between genetic and phenotypic correlations (Table 3.2).

Table 3.2: Genetic and phenotypic correlations among protein, oil and seed size in F6-derived soybean population from N87-984-16 x TN93-99 averaged over six environments (2002 and 2003).

Trait	Protein	Oil	Seed weight
Protein	-0.60***	-0.59***	0.50***
Oil	0.54***		-0.03 ^{ns}
Seed weight		-0.05 ^{ns}	

Note: Genetic correlations are in the lower left diagonal whereas phenotypic correlations are in the upper right diagonal of the table.

*** = Significant at 0.001 probability level; ns = non-significant ($P > 0.05$).

A strong negative phenotypic correlation between protein and oil ($r = -0.84$, $P < 0.001$) was reported recently (Chung et al., 2003) and it was much stronger ($r = -0.98$, $P < 0.001$) in a different study (Mansur et al., 1996). On the basis of 50 year average data, a moderately weak negative correlation between protein and oil ($r = -0.39$, $P < 0.001$) was reported and it was slightly weaker ($r = -0.33$, $P < 0.001$) in the northern and stronger ($r = -0.52$, $P < 0.001$) in the Southern states of USA (Yaklich et al., 2002). Yaklich et al. (2002) have also reported the correlation coefficient on the basis of maturity groups 00 to VIII and showed that there was a stronger correlation as the maturity group increases. Plant breeders prefer to have the correlation between protein and oil in soybean as weak as possible to enable flexible targeting of phenotype.

QTL analyses for seed traits

QTL for protein

There were two QTL linked to molecular markers Satt274 and Satt570 associated with protein concentration in soybean in this population, distributed in two molecular linkage groups D1b and G respectively (Table 3.3). The QTL near Satt570 was consistent in all the environments. When LOD scores from combined data were plotted, there was a peak above Satt570 (Figure 3.4), further indicating that the genomic region near Satt570 has a strong QTL associated with seed protein concentration.

Table 3.3: Quantitative trait loci, map position and genetic contribution for protein, oil and seed size in an F6-derived soybean population from N87-984-16 x TN93-99.

Trait	Environment†	LG‡	Marker	Position (cM)§	LOD¶ Score	Additive Effect#	R ² (%)
Protein	Ames 2002	D1b	Satt274	116.35	2.8	0.30	24.9
	Holston 2003	G	Satt570	12.74	3.5	3.80	15.4
	Holston	G	Satt570	12.74	2.6	0.33	12.8
	KPSF 2002	G	Satt570	12.74	2.6	2.88	13.2
	KPSF	G	Satt570	12.74	3.3	3.30	15.0
	Combined 2002	G	Satt570	12.74	2.7	0.20	14.1
	Combined	G	Satt570	12.74	3.5	0.31	20.2
Oil	Ames 2003	D1b	Satt274	116.35	3.4	1.94	13.1
	Holston 2003	D1b	Satt274	116.35	3.3	1.97	13.8
	KPSF 2002	D1b	Satt274	116.35	3.8	2.32	13.8
	KPSF	D1b	Satt274	116.35	3.4	2.07	11.8
	Combined 2003	D1b	Satt274	116.35	3.0	1.84	11.8
	Ames 2002	G	Satt570	12.74	3.3	-0.16	13.2
	Ames	G	Satt570	12.74	3.2	-0.12	11.4
	KPSF 2002	G	Satt570	12.74	3.0	-2.31	14.1
	KPSF	G	Satt570	12.74	3.4	-2.41	16.8
	KPSF 2003	H	Satt293	89.09	3.6	-2.73	18.1
	KPSF	H	Satt317	89.52	2.8	-1.84	9.7
	Combined 2003	H	Satt317	89.52	2.6	-1.62	9.4
	Ames 2002	I	Satt562	22.84	3.3	-0.18	15.9
	Ames 2003	O	Satt420	49.71	2.6	1.68	10.5
	Combined 2002	O	Satt420	49.71	3.5	0.15	15.0
	Holston 2002	O	Satt479	54.20	3.5	0.19	13.2
	Combined	O	Satt479	54.20	3.1	0.14	12.0
Seed size	Combined 2002	D1a	Satt147	108.89	4.6	-3.6	16.5
	Holston 2002	D1a	Satt147	108.89	4.2	-0.5	18.1
	Holston 2003	D1a	Satt147	108.89	3.9	-0.5	14.7
	KPSF 2002	D1a	Satt147	108.89	3.1	-0.4	11.8
	Combined 2002	D1a	Satt184	17.52	4.8	-4.9	29.9
	Holston 2002	D1a	Satt184	17.52	4.8	-0.7	37.8
	Combined	D1a	Satt184	17.52	3.5	-3.1	11.3
	KPSF 2002	D1a	Satt184	17.52	3.6	-0.8	39.9
	Holston 2002	D1b	Satt459	118.62	3.2	-0.4	10.5
	KPSF 2003	D1b	Satt459	118.62	3.2	-0.5	14.6
	Combined 2002	D2	Satt002	47.73	3.5	-3.2	11.9
	Combined	D2	Satt002	47.73	2.9	-3.0	10.0
	KPSF 2002	D2	Satt002	47.73	3.4	-0.5	11.8
	Holston 2003	E	Satt185	44.76	4.1	-0.5	13.5
	Holston 2003	E	Satt268	44.27	3.0	-0.4	10.4
	Holston 2002	M	Satt540	35.85	4.0	0.5	16.7
	Ames 2003	N	Satt255	76.49	3.0	3.9	15.0

† A total of six environments were utilized in 2002 and 2003. Ames Plantation near Grand Junction (Ames), TN; the Holston unit of the Knoxville Experiment Station in Knoxville, TN (Holston), and the Knoxville Plant Science Farm (KPSF) at the Knoxville Experiment Station in Knoxville, TN.; ‡ Linkage group; § This position is based on the composite integrated genetic linkage map of the soybean; ¶ Log of odds ratio, # ½(P1-P2)

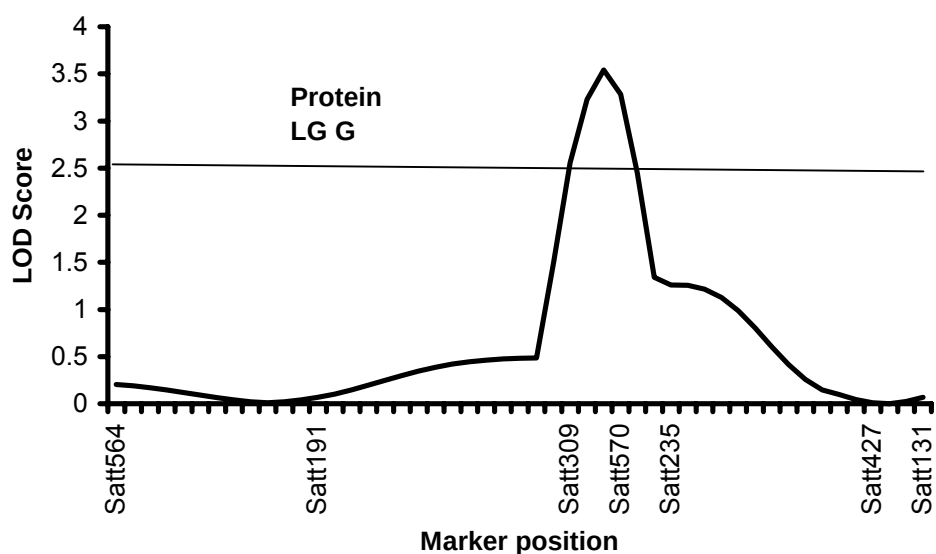


Figure 3.4: LOD score plot for protein concentration on LG G in combined data. There is a major QTL on LG G near Satt570.

The other QTL near Satt274 was associated with protein in only one environment. When its LOD scores are plotted on LG D1b, there was a peak above Satt274 (Figure 3.5). From these results, one major QTL near Satt570 appears to be stable for protein concentration in this population whereas the QTL near Satt274 was environmentally sensitive. Both of these QTL (Satt274 and Satt570) were found associated with seed nitrogen accumulation during reproductive growth stages in soybean (Panthee et al., 2004).

Nitrogen accumulated in the soybean seed is involved in storage protein accumulation. For both the protein QTL, favorable alleles were contributed through P1 (N87-984-16). P1 is the high parent for protein content.

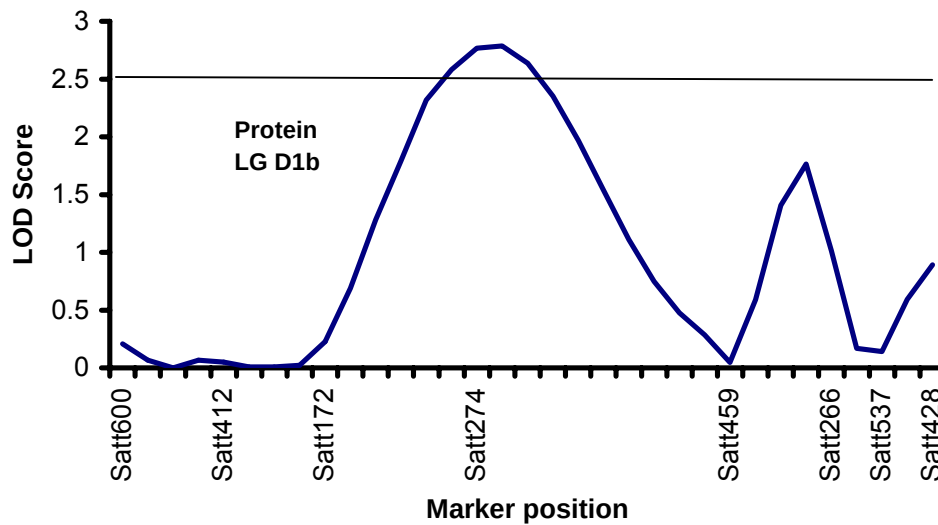


Figure 3.5: LOD score plot for protein concentration on LG D1b in combined data averaged over six environments (2002 and 2003). There is a major QTL on LG D1b Satt274.

Phenotypic variation explained by an individual QTL ranged from 12.8 to 24.9% (Table 3.3). From phenotypic point of view, Satt274 was the more important locus since it explained most of the variation. However, Satt570 linked QTL had the greatest additive genetic effect in the population for protein content. This information is important knowledge for allele substitution in breeding programs. In the case of Satt274 as well as Satt570, we need to substitute the allele through P1 to target genetic gains for increased seed protein concentration.

Recently, genomic region linked to Satt459 has been reported as a novel protein QTL (Hyten et al., 2004), which is within 2.3 cM downstream of

Satt274, a QTL identified in our population. This indicates that our group has verified a new genomic region controlling protein concentration in soybean. In another study, 14 QTL were found to be associated with protein concentrations distributed among nine different linkage groups (Brummer et al., 1997). Interestingly, they found three QTL associated with protein on linkage group G, and we also found a major stable QTL near Satt570 in this linkage group. However, the closest QTL from the Brummer et al. (1997) study is located at least 54.8 cM distal to Satt570, indicating that we have detected a novel protein QTL from the Prolina parentage. These findings suggest that out of two protein QTL, the one near Satt274 is in a genomic region recently identified whereas the one near Satt570 is novel.

QTL for oil

There were seven QTL linked to molecular markers Satt274, Satt293, Satt317, Satt420, Satt479, Satt562 and Satt570 associated with oil concentration in this population (Table 3.3). These QTL were distributed among five molecular linkage groups D1b, G, H, I and O. Phenotypic variation explained by an individual QTL ranged from 9.4 to 18.1%. Favorable alleles for increased oil content in three of the seven QTL linked to Satt274, Satt420 and Satt479 were contributed through P1, although P2 was the high parent for oil content. It is noted that Satt293 and Satt317 on LG H are just 0.43 cM apart on the composite integrated molecular linkage

map (Song et al., 2004), thus representing virtually the same genomic region. Similarly, Satt420 and Satt479 on LG O are located 4.5 cM apart indicating that this is another genomic region associated with oil content in soybean.

One QTL linked to RFLP marker B072_1 was found to be associated with oil content in a soybean population from 'Peking' x 'Essex' (Qiu et al., 1999), which is about 20 cM downstream of Satt293 and Satt317 from our population in the composite integrated linkage map (Song et al., 2004). The distance between the regions identified by Qiu et al. (Qiu et al., 1999) on LG H and the region on that same LG that we detected is still wide, but provides additional information on oil controlling region in soybean. At least 12 QTL were detected associated with oil content in a study distributed over seven linkage groups (Brummer et al., 1997). In that study, only one QTL near A069_1 was found to be located on linkage group H; we also found a QTL on that linkage group but it was about 56 cM distant to A069_1, indicating independent QTL. Another QTL near RFLP marker A566_2 located on LG H was identified (Lee et al., 1996). It has been recently reported that there are three genomic regions (Pantalone et al., 2004) containing QTL associated with oil concentration on LG H, all of them are at least 30 cM apart from the QTL detected in the present study indicating that our population may contain novel QTL associated with oil concentration.

Three QTL associated with oil content from LG G were previously identified (Lee et al., 1996). Two of them lined to L154_2 and A235_1 were more than 85 cM apart from the one detected in the present population, indicating that the oil QTL we detected was independent. Similarly, oil QTL linked to L002_1 detected by Lee et al. (1996) was 39 cM distal to QTL near Satt570, indicating that this is also an independent QTL. Satt562 detected in our study was previously detected by Csanadi et al. (2001), thus our finding verifies that oil locus.

There are no reported QTL from molecular linkage group D1b and O associated with oil concentration (Pantalone et al., 2004). Therefore, QTL reported from the present study on those linkage groups are novel.

Two populations were studied for QTL analyses for oil and protein concentration by Lee et al. (1996), however no QTL were consistent between the populations. Recently, Fasoula et al. (2004) reported that only two of four previously reported protein QTL and two of three previously reported oil QTL could be confirmed in a population of 'PI97100' x 'Coker 237'. In the same study, they used a different population of 'Young' x 'PI416937', in which they could not confirm any of the previously reported three protein QTL, and only one of three oil QTL was confirmed. Therefore, soybean geneticists should continue efforts to

detect, validate and confirm QTL to make this technology more practical for applied breeders.

It was interesting to note that both the protein QTL detected in the present study were also associated with oil content (Table 3.3). However, one of those two QTL near Satt570 had opposite additive genetic effects for these two traits whereas that near Satt274 had positive additive genetic effects on both traits. In the present population, there was a moderate negative correlation between protein and oil content (Table 3.2). Possibly the region near Satt274 is playing role for such a reduced negative correlation, otherwise a strong negative correlation may have been observed between protein and oil concentrations. To increase the protein and oil content simultaneously in a soybean line, the most desirable condition is having positive effects for both traits on particular loci. Therefore, QTL linked to Satt274 may be useful for this purpose. Three QTL linked to T155, A329 and Satt006 located on LG A1, B2 and L respectively in the integrated linkage map were associated with protein as well as oil content in soybean population (Mansur et al., 1996). However, all three loci had opposite effects on protein and oil content making it difficult to improve genetic gain in both traits simultaneously using these markers.

QTL for seed size

There were eight QTL linked to molecular markers Satt002, Satt147, Satt184, Satt185, Satt255, Satt268, Satt459 and Satt540 associated with seed size in this population distributed among six molecular linkage groups D1a, D1b, D2, E, M and N (Table 3.3). Four of these QTL near Satt002, Satt147, Satt184 and Satt459 showed significant associations with seed size in more than one environment. Thus, they are regarded as stable QTL for seed size in this population. The remaining four QTL near Satt185, Satt255, Satt268 and Satt540 were associated with seed size in only one environment, thus they are environmentally sensitive (Brummer et al., 1997). Phenotypic variation explained by an individual QTL ranged from 10 to 39.9% (Table 3.3).

Favorable alleles for seed size were found at six loci linked to Satt002, Satt147, Satt184, Satt185, Satt268 and Satt459 contributed through P2, although P1 was the high parent for seed size. Two loci near Satt255 and Satt540 received the alleles for increased seed size through P1. In terms of genetic contribution, locus near Satt184 (located on LG D1a) had the most additive genetic effect on seed size in this population (Table 3.3). Linkage group D1a and E contain two seed size QTL each (loci near Satt147 and Satt184 are within 20 cM distance on LG D1a, and loci near Satt185 and Satt268 are within 1.5 cM distance on LG E).

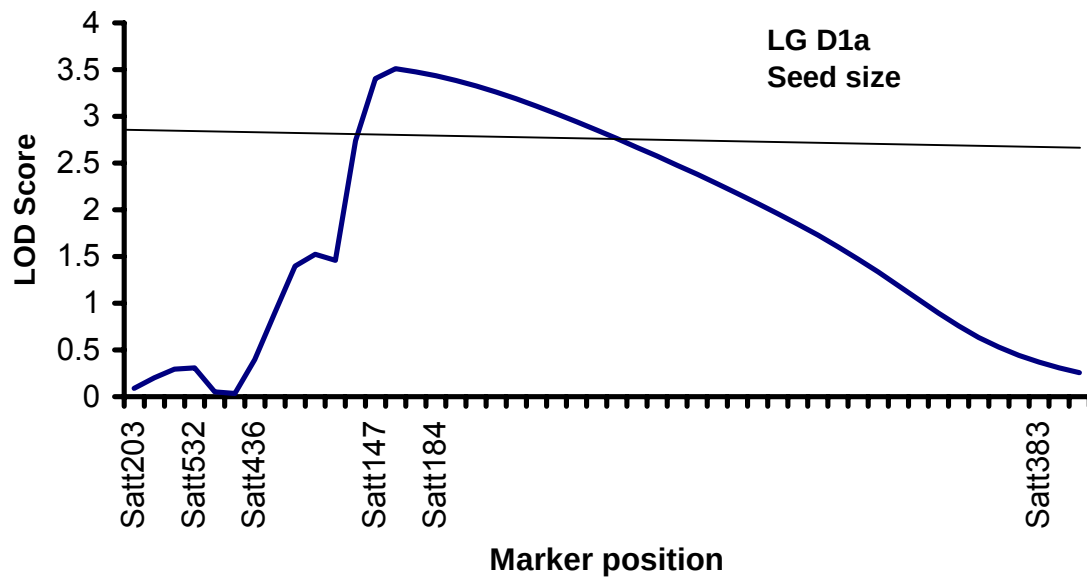


Figure 3.6: LOD score plot for seed size on LG D1a in combined data averaged over six environments (2002 and 2003). There is a major QTL flanked by Satt147 and Satt184 on LG D1a.

We plotted the LOD score against markers on LG D1a, which revealed that there was a QTL between Satt147 and Satt184 in this population (Figure 3.6). A relatively flat peak indicated that there could be more than one major QTL in this genomic region. The locus near Satt540 detected in our population was also reported to be significantly associated with seed size in a population derived from 'Essex' x 'Williams' (Chapman et al., 2003) and in another population (Csanadi et al., 2001).

In a different study, seven QTL were identified to be associated with seed size on the basis of single factor ANOVA (Lee et al., 2001) on the basis of RFLP markers. They identified QTL near A095 on LG D2, and we

identified QTL near Satt002 on the same linkage group. However, these two QTL are about 44.2 cM apart in the integrated linkage map (Song et al., 2004). Thus, they may represent different seed size loci. Lee et al. (2001) also found QTL near A069, located on LG E, and we found QTL near Satt185 and Satt268 in our study, which are near to those RFLP markers.

Three populations were developed from six parents to study QTL associated with seed size by Hoeck et al. (2003). They grew all the populations in four environments and mapped QTL in an individual environment as well as in a combined data set. They found as many as 22 QTL in one population in an environment. Interestingly, there were very few common QTL among the three populations of their study, yet there were two QTL near Satt002 and Satt184 associated with seed size common with our population. These two QTL appeared to be stable in our population (Table 3.3). Fasoula et al. (2004) recently confirmed seed size QTL in independent populations and wisely mentioned that it is necessary to verify already reported QTL in independent populations. Our study has verified some of the previously reported seed size QTL and has reported new QTL for that trait. The QTL near Satt002, Satt184 and Satt540 are verified QTL for seed size, whereas those near Satt147, Satt255 and Satt459 are novel.

QTL associated with seed size were mapped in two populations in a different study (Mian et al., 1996). Seven QTL were found associated with seed size in one population and nine in others. Although QTL across the environments were consistent, there were only two QTL from the same genomic regions consistent across the populations. None of these were detected in our population.

Interestingly, almost all the QTL detected for protein, oil and seed size in our population had $R^2 \geq 10\%$, thus they are major QTL for these seed traits (Falconer and Mackay, 1996). Analyses of allelic contributions for quantitative traits at particular loci are important to improve traits through introgression of alleles from a particular breeding line. In a population developed from *G. max* x *G. soja*, for the study of QTL associated with protein and oil, it was found that all protein related alleles were contributed through *G. soja* whereas oil related QTL were contributed through *G. max* (Diers et al., 1992). However, in soybean populations such as ours, we found contribution from both parents. Similar observations were made for a number of traits in another study (Mansur et al., 1993). Transgressive segregation may take place among the progeny, and molecular genetic detection of the underlying loci, which contribute to the process is important knowledge for plant breeders.

On the basis of results from past experiments and present findings, it is important to note that all QTL associated with a trait are not stable across populations and environments. In MAS, it would be more desirable to use confirmed QTL. For this purpose, it is necessary to conduct QTL mapping studies in as many and as diverse environments as feasible (Fasoula et al., 2004; Song et al., 2004). The present experiment provides important new information and we encourage other researchers to continue working in this direction.

Finally, our population served not only for QTL detection but also was utilized to develop and release new improved protein germplasms TN04-5321 by the Tennessee Agricultural Experiment Station. We hope that the QTL identified from this study and the recent germplasm that we have released will benefit breeders in accumulating favorable alleles for improvement in soybean seed quality.

References

- Alt, B.J., W.R. Fehr, and G.A. Welke. 2002. Selection for large seed and high protein in two- and three-parent soybean populations. *Crop Sci* 42:1876-1881.
- Brim, C.A. 1966. A modified pedigree method of selection in soybeans. *Crop Sci* 6:220.

- 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. 1987. Quantitative genetics: Results relevant to soybean breeding, p. 211-247, *In* J. R. Wilcox, ed. *Soybeans: Improvement, Production and Uses*, Second ed. American Society of Agronomy, Inc., Madison, WI, USA.
- Burton, J.W., T.E. Carter, and R.F. Wilson. 1999. Registration of Prolina soybean. *Crop Sci* 39:294-295.
- Chapman, A., V.R. Pantalone, A. Ustun, F.L. Allen, D. Landau Ellis, R.N. Trigiano, and P.M. Gresshoff. 2003. Quantitative trait loci for agronomic and seed quality traits in an F_2 and $F_{4:6}$ soybean population. *Euphytica* 129:387-393.
- Chung, J., H.L. Babka, G.L. Graef, P.E. Staswick, D.J. Lee, P.B. Cregan, R.C. Shoemaker, and J.E. Specht. 2003. The seed protein, oil, and yield QTL on soybean linkage group I. *Crop Sci* 43:1053-1067.
- Churchill, G.A., and R.W. Doerge. 1994. Empirical threshold values for quantitative trait mapping. *Genetics* 138:963-971.
- Cregan, P.B., T. Jarvik, A.L. Bush, R.C. Shoemaker, K.G. Lark, A.L. Kahler, N. Kaya, T.T. VanToai, D.G. Lohnes, and J. Chung. 1999. An integrated genetic linkage map of the soybean genome. *Crop Sci* 39:1464-1490.

- Csanadi, G., J. Vollmann, G. Stift, and T. Lelley. 2001. Seed quality QTLs identified in a molecular map of early maturing soybean. *Theor Appl Genet* 103:912-919.
- 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.
- Falconer, D.S., and T.F.C. Mackay. 1996. *Introduction to Quantitative Genetics*. Fourth ed. Longman Group Ltd.
- Fasoula, V., D. Harris, and H. Boerma. 2004. Validation and designation of quantitative trait loci for seed protein, seed oil, and seed weight from two soybean populations. *Crop Sci* 44:1218-1225.
- Hoeck, J.A., W.R. Fehr, R.C. Shoemaker, G.A. Welke, S.L. Johnson, and S.R. Cianzio. 2003. Molecular marker analysis of seed size in soybean. *Crop Sci* 43:68-74.
- Hyten, D.L., V.R. Pantalone, C.E. Sams, A.M. Saxton, D. Landau-Ellis, T.R. Stefaniak, and M.E. Schmidt. 2004. Seed quality QTL in a prominent soybean population. *Theor Appl Genet* 109:552-561.
- Kearsey, M.J., and H.S. Pooni. 1996. *The Genetical Analysis of Quantitative Traits*. First ed. Chapman and Hall, New York.
- Lee, S.H., K.Y. Park, H.S. Lee, E.H. Park, and H.R. Boerma. 2001. Genetic mapping of QTLs conditioning soybean sprout yield and quality. *Theor Appl Genet* 103:702-709.

- Lee, S.H., M.A. Bailey, M.A.R. Mian, T.E. Carter, Jr., E.R. Shipe, D.A. Ashley, W.A. Parrott, R.S. Hussey, and H.R. Boerma. 1996. RFLP loci associated with soybean seed protein and oil content across populations and locations.
- Lincoln, S.E., M.J. Daly and E.S. Lander. 1993. Constructing genetic linkage maps with MAPMAKER Version 3.: A tutorial and reference manual. Release 3rd edition. Whitehead Institute for Biomedical Research, Cambridge, MA.
- Liu, K. 1997. Soybeans: chemistry, technology and utilization Chapman and Hall.
- 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.
- Mansur, L.M., J.H. Orf, K. Chase, T. Jarvik, P.B. Cregan, and K.G. Lark. 1996. Genetic mapping of agronomic traits using recombinant inbred lines of soybean. *Crop Sci* 36:1327-1336.
- Maughan, P.J., M.A. Saghai Maroof, and G.R. Buss. 1996. Molecular-marker analysis of seed-weight: genomic locations, gene action, and evidence for orthologous evolution among three legume species. *Theor Appl Genet* 93:574-579.
- 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* 93:1011-1016.
- Nyquist, W.E. 1991. Estimation of heritability and prediction of selection response in plant populations. *Crit Rev Plant Sci* 10:235-322.
- Orf, J.H., K. Chase, T. Jarvik, L.M. Mansur, P.B. Cregan, F.R. Adler, and K.G. Lark. 1999. Genetics of soybean agronomic traits. I. Comparison of three related recombinant inbred populations. *Crop Sci* 39:1642-1651.
- Pantalone, V., D. Walker, R. Dewey, and I. Rajkan. 2004. DNA marker-assisted selection for improvement of soybean oil concentration and quality, p. 283-311, *In* R. Wilson, et al., eds. *Legume crop genomics*. AOCS Press, Champaign, IL.
- Pantalone, V.R., F.L. Allen, and D. Landau-Ellis. 2003. Registration of 'TN93-99' Soybean germplasm. *Crop Sci* 43:1137.
- Panthee, D.R., V.R. Pantalone, C.E. Sams, A.M. Saxton, D.R. West, and W.E. Rayford. 2004. Genomic regions governing soybean seed nitrogen accumulation. *J Am Oil Chem Society* 81:77-81.
- Qiu, B.X., P.R. Arelli, and D.A. Sleper. 1999. RFLP markers associated with soybean cyst nematode resistance and seed composition in a 'Peking' x 'Essex' population. *Theor Appl Genet* 98:356-364.
- SAS Institute Inc. 2002. SAS guide for personal computers. 9th Edition ed. SAS Institute, Cary, NC.

- Song, Q.J., L.F. Marek, R.C. Shoemaker, K.G. Lark, V.C. Concibido, X. Delannay, J.E. Specht, and P.B. Cregan. 2004. A new integrated genetic linkage map of the soybean. *Theor Appl Genet* 109:122 - 128.
- Wang, S, C, .J. Basten and Z.B. Zeng 2003. Windows QTL Cartographer. Release V2.0. North Carolina State University, Raleigh, NC.
- Wilson, R.F. 2004. Seed composition, *In* H. R. Boerma and J. E. Specht, eds. *Soybeans: Improvement, Production and Uses*, 3rd Edition ed. ASA, CSSA, SSA, Madison, WI.
- Yaklich, R.W., B. Vinyard, M. Camp, and S. Douglass. 2002. Analysis of seed protein and oil from soybean Northern and Southern Region Uniform Tests. *Crop Sci* 42:1504-1515.

PART 4

Quantitative Trait Loci for β -Conglycinin (7S) and Glycinin (11S)

Fractions of Soybean Storage Protein

This chapter is a lightly revised version of paper by the same name that has been published in ***Journal of American Oil Chemists' Society*** **2004 81: 1005-1012** by D.R. Panthee, P. Kwanyuen, C.E. Sams, D.R. West, A.M. Saxton and V.R. Pantalone

My use of “our” in this chapter refers to my co-authors and myself. My primary contribution to this paper include (1) organization, development and management of all aspects of the project, (2) collecting all molecular data using polymorphic molecular markers, (3) collecting all the field data at Knoxville environments, (4) analyzing for protein fractions, (5) analyzing all data, and (6) all of the writing except for some minor revisions.

Abstract

Glycinin (11S) and β -conglycinin (7S) are important seed storage proteins in soybean [*Glycine max* (L.) Merr.]. A major limitation of soybean seed storage proteins is their low levels of the sulfur-containing amino acids, methionine and cysteine, which are important nutritional components of protein meal. Glycinin contains significantly more S-containing amino acids than does β -conglycinin. Thus detection of quantitative trait loci (QTL), which govern 11S, may provide marker assisted selection (MAS) opportunities to improve soybean total S-containing amino acids. The objective of this study was to detect and map QTL governing 7S and 11S fractions of soybean seed storage proteins. To achieve this objective, 101

F₆-derived recombinant inbred lines (RIL) developed from a cross of N87-984-16 x TN93-99 were used. Storage proteins were extracted from all RIL and separated in 10-20% linear gradient polyacrylamide gels. Dried gels were scanned for individual sub-units of storage protein with a densitometer equipped with a He-Ne laser light source. Data were converted to concentration for each sub-unit component, and analyzed using SAS software. A significant ($P<0.05$) difference among genotypes was found for glycinin and β -conglycinin. A total of 94 polymorphic simple sequence repeat (SSR) molecular genetic markers were utilized in screening all RIL. Three QTL for glycinin linked to Satt461, Satt292 and Satt156 were located on LG D2, I and L respectively, whereas two QTL for conglycinin linked to Satt461 and Satt249 were located on LG D2 and J. Phenotypic variation explained by individual QTL ranged from 9.5 to 22%. These QTL may provide useful marker assisted selection opportunities for improvement of nutritional quality in soybean.

KEY WORDS β -conglycinin, *Glycine max*, glycinin, QTL mapping, soybean, storage protein.

Introduction

Soybean is a major source of protein and amino acids for human and animal feeds in the world (Clarke and Wiseman, 2000). However, soy protein is not perfect because of its low levels of the sulfur-containing essential amino acids, methionine and cysteine. Enhancing the

concentration of nutritionally essential amino acids in soybean would improve soy meal quality and boost its inclusion in diets for humans and animals.

Soybean storage protein has two major fractions, β -conglycinin (7S) and glycinin (11S) accounting for more than 70% of the total proteins (Tsukada et al., 1986). β -conglycinin is a trimer with sub-units α , α' and β , and a molecular weight of about 180 kDa. Glycinin is a hexamer with a molecular weight of 360 kDa, consisting of acidic and basic sub-units. Glycinin comprises about 35% of the total seed storage protein (Yagasaki et al., 1997). Glycinin has 3 to 4 times more S-containing amino acids (particularly methionine) than does β -conglycinin (Kitamura, 1995). In addition, the β sub-unit is known to be void of methionine and cysteine. Hence the glycinin fraction may be more desirable than the β -conglycinin fraction in developing improved amino acid balance in soy meal. Both 7S and 11S are components of total seed protein, and there is a typical inverse relationship between 7S and 11S concentrations (Ogawa et al., 1989). Thus, glycinin can be increased at the expense of β -conglycinin. Mutants with increased glycinin concentration have been generated using this concept by some plant breeders (e.g. Kitagawa et al., 1991; Takahashi et al., 1996).

Yaklich (2001) studied the storage protein of high protein lines of soybean to determine how the two major storage proteins contribute to the increased protein content. It was found that high seed protein lines contain more β -conglycinin and glycinin than normal protein lines. These data also inferred that quantitative differences exist in protein bands of major seed storage proteins, indicating that there should be differences in the genome that are responsible for the synthesis of different quantities of sub-units of the storage protein. In fact, Nielsen et al. (1989) identified five genes regulating the production of glycinin in soybean. Harada et al. (Harada et al., 1989) showed that there are three genes, which control the production of β -conglycinin. However, the position of those genes on the soybean genetic linkage map remains to be determined.

Hayashi et al. (2000) used AFLP markers to locate the *cgdef* locus in soybean. This locus determines the β -conglycinin content in soybean, and a single recessive gene reduces the β -conglycinin content. Such reduction results in an increased level of glycinin. A mutant line (Keburi) with reduced β -conglycinin was also identified by Kitamura and Kaizuma (1981).

Among the efforts to increase the S-containing amino acids in the soy protein fraction, genetic transformation was accomplished by introducing a gene from Brazil nut (*Bertholletia excelsa*), but the project was abandoned

because of allergenicity issues resulting from the Brazil nut transgene (Mandal and Mandal, 2000). *Glycine soja*, the wild ancestor of soybean, has also been studied to increase glycinin, however, hybridizations between soybean and *G. soja* have not yet led to the development of an agriculturally acceptable cultivar (Kwanyuen et al., 1997; Liu, 1997).

In our experiment, the parental line N87-984-16 is one of two essentially equivalent F_8 -derived sister lines that constitute the high protein commercial soybean cultivar 'Prolina' (Burton et al., 1999). Prolina has high seed protein concentration and also exhibits superior protein gelling properties, indicating that its storage proteins, may contain more disulfide bonds and increased S-containing amino acids (Luck et al., 2001). The other parental line utilized in our experiment, TN93-99, is a high yielding soybean line with typical protein concentration. TN93-99 is the registered germplasm GP-280 (Pantalone et al., 2003). The objective of the study was to detect and map QTL governing 7S and 11S fractions of the storage proteins in a soybean RIL population formed from the cross between N87-984-16 x TN93-99.

Materials and Methods

Plant material

A total of 101 F_6 -derived recombinant inbred lines (RIL) were developed from a cross of N87-984-16 x TN93-99. The N87-984-16 line is one of two

sister lines, whose blend constitutes the high protein commercial cultivar 'Prolina' (Burton et al., 1999). The line TN93-99 is a high yielding and well adapted University of Tennessee breeding line and is currently registered as a germplasm (Pantalone et al., 2003). Crossing was made during summer of 1998. F₁ seeds were harvested in October 1998 and F₂ single plants were grown in Costa Rica during winter 1998. Generations were advanced until F₅ in Costa Rica and F_{5:6} seeds were obtained in May 2000. About 300 F_{5:6} single plants were grown at Knoxville as a Recombinant Inbred Line (RIL) population and 101 random plants of similar Tennessee adapted maturity were selected as a source population for this study. A total of 101 F₆-derived recombinant inbred lines developed from a cross of N87-984-16 x TN93-99 were planted in a randomized complete block design (RCBD) with three replications at the Knoxville Experiment Station of the University of Tennessee, Knoxville, TN in May 2001 and 2003. Each line was planted in plots of 6m length with 75cm between rows. In 2001, two-row plots were used, and four-row plots were used in 2003. Seeds were harvested with a two-row combine and stored below 10°C until analyzed.

Preparation of protein

Ten grams of soybean seeds were ground in a water-cooled (20°C) Knifetec 1095 Sample Mill (FOSS Tecator, S-26321, Hogana, Sweden) for 20 seconds. This setting produced soybean flour with a uniform particle

size. Soluble protein was extracted for 1 h at room temperature while stirring from one gram of full-fat soybean flour in a 1:15 (w/v) ratio with 0.2M Tris-HCl buffer, pH 8.0, that contained 0.1M β -mercaptoethanol. The mixture was centrifuged at 10,000g for 10 min at 4°C. After the fat layer was removed, an aliquot of 1mL of crude protein extract or supernatant was taken from each sample. The total protein concentration of each sample was determined as described by Bradford (1976). The modified protein assay kit was obtained from Bio-Rad (Richmond, CA). Storage proteins and their polypeptides were dissociated in the crude extract by adding an equal volume of 5% SDS and 0.1M β -mercaptoethanol solution each to each sample and boiling in a water bath for 10 minutes for complete dissociation. Glycerol and bromophenol blue were added to each sample to achieve a final concentration of 10% and 0.025% respectively.

Polyacrylamide protein gel electrophoresis

Proteins and their polypeptides were separated in a Bio-Rad (Richmond, CA) Protean II vertical slab gel apparatus according to Chua (1980) with the following modifications. Each sample containing approximately 80 to 100 μ g protein (approximately 10 μ L of protein sample) was loaded onto the gel. Proteins and polypeptides were separated using a linear gradient of 10 to 20% polyacrylamide gel. The dimensions of the separating gel are 14 x 16 x 0.15cm³ with 15 sample wells in the stacking gel. Blank sample

wells were left between loaded samples to prevent protein cross-contamination during electrophoresis and to facilitate accurate quantitation by scanning densitometry after electrophoresis. Since more than one gel was required for these protein samples, and since it was necessary to obtain the same polyacrylamide composition throughout the entire experiment, identical gradient gels were cast from the same polyacrylamide solution and from the same gradient-producing condition. Electrophoresis of each protein sample was carried out in duplicate at a constant current of 10mA/gel at room temperature until the bromo-phenol blue or tracking dye reached the bottom of the gel.

Staining and destaining

Gels were fixed in 40% (v/v) methanol and 10% (v/v) acetic acid for at least half an hour prior to staining. They were stained in freshly prepared dye containing 0.25% Coomassie Brilliant Blue (w/v), 40% (v/v) methanol and 10% (v/v) acetic acid on an orbit shaker with fixed speed. They were stained for at least 8 h. Gels were then destained in 40% (v/v) methanol and 10% (v/v) acetic acid on the same orbital shaker. The destaining solution was changed every two hours for a minimum of four changes. The destaining was terminated when the background gels were almost visibly cleared of the dye. The residual dye on the gels did not interfere with the analysis of polypeptide bands since local average background subtraction was used in the analyses. Destained gels were soaked in

deionized water for at least 5 min. Each gel was placed between two cellophane sheets and dried in a Bio-Rad (Richmond, CA) GelAir dryer.

Scanning densitometry. Dried gels were scanned as described by Kwanyuen and Wilson (2000) with a Molecular Dynamics Personal Densitometer SI (Sunnyvale, CA) equipped with a HeNe laser light source. ImageQuant (Molecular Dynamics, Sunnyvale, CA) software for volume integration was used in data analysis to determine total absorbance of entire protein bands (Figure 4.1). The absorbance of protein on gels was adjusted to ensure that the most intense protein bands were within the optimal and linear response range of the detector. Apparent absorbance of each protein was obtained by subtracting the background absorbance from the total absorbance of the protein bands within the same gel volume. The relative amount of each protein or polypeptide was expressed as a percentage of total protein in the same gel lane.

DNA extraction and Polymerase Chain Reaction

DNA was extracted from the RIL and parental lines utilizing Qiagen Plant Easy DNA Extraction Kit (Qiagen Huden, Germany). Polymerase chain reaction (PCR) consisted of 7.4 μL of ddH₂O, 1 μL of 10X PCR Buffer, 1 μL of 2 mM dNTPs mixture (Pharmacia, Piscataway, NJ), 0.5 μL of 20 μM forward and reverse primer, 0.1 μL of 5 units/ μL Klentaq (Ab Peptides Inc., St. Louis, MO) and 2 μL of 20 ng/ μL template DNA. The PCR was

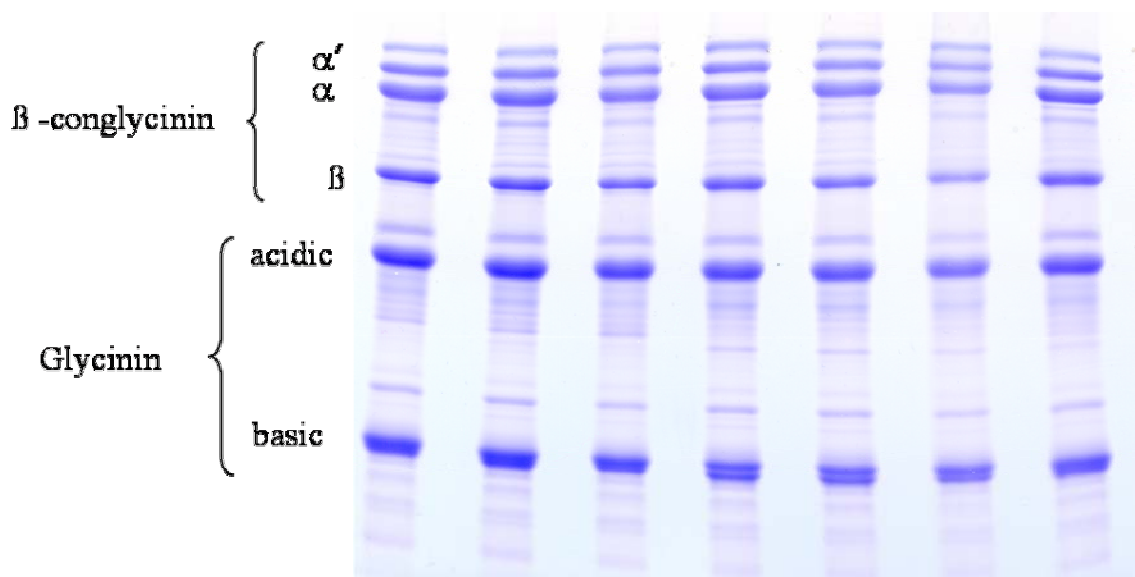


Figure 4.1: Banding pattern of the components of 7S (α' , α and β) and 11S (acidic and basic) fraction of soy protein in SDS-PAGE obtained from soybean RIL of N87-984-16 x TN93-99.

performed in a 96-well MBS Hybaid thermocycler (Hybaid, Franklin, MA). PCR conditions were a) 94° C for five minutes followed by 35 cycles at 94° C for denaturation for 25 s, b) 47° C for annealing for 30 s, c) 72° C for 25 s for extension, and one last cycle at 72° C for final extension for five minutes. Parents were screened for a total of 550 (ATT)_n type of simple sequence repeat (SSR) genetic markers developed by Cregan *et al.* (1999). The sequence information for the markers is publicly available from the Website (www.soybase.org), which was verified on December 16, 2004. A total of 94 SSR markers, which were found to be polymorphic in the parents and among the RIL, were screened and used in QTL analyses.

DNA gel electrophoresis

A 6% non-denaturing polyacrylamide gel electrophoresis (PAGE) used to separate the PCR product consisted of 6% bis-acrylamide, 0.5% TBE buffer, 0.07% APS and 0.035% TEMED, which was measured as 28.5 mL of 40% bis-acrylamide, 160.2 mL of 0.593X TBE buffer, 1.33 mL of 10% APS and 66.5 μ L of TEMED. Two μ L of loading buffer (6X) were added to PCR products and a 10 μ L sample was loaded on the gel. The running buffer was 0.5X TBE. The gel was run at constant 300V for 3 h. Circulated air was used to keep the glass plates cool during gel running. Ethidium bromide (50 μ L, 10 mg/mL) was added to the running buffer in order to visualize the bands under exposure to UV light. Bands were scored using 1 to represent P_1 (N87-984-16), 2 to represent heterozygote, and 3 to represent P_2 (TN93-99) alleles for each primer locus.

Data analysis

The storage proteins (phenotypic data) β -conglycinin and glycinin, and their sub-units α' , α , β , acidic and basic data from 2001 and 2003 were analyzed separately as well as combined using SAS software (SAS Institute Inc, 2000). Association between molecular marker and phenotype was determined by single factor ANOVA utilizing PROC GLM of SAS using least square means from 2001, 2003 and combined data. With this analysis, molecular markers linked to QTL for storage protein and its components were identified. We confirmed the QTL by composite interval

mapping (CIM) of QTL Cartographer version 2.0 using model 6 (standard model) with 10 cM window size, 2 cM walking speed and five control molecular markers.

Results and Discussion

The β -conglycinin (7S) content was derived by summation of the original scanned value of Alpha prime (α'), Alpha (α) and Beta (β), and the glycinin (11S) content was derived by summation of the acidic and basic components (Figure 4.1). The values obtained from the densitometer scans were converted to percentage of the total protein in the lane. When selecting a genotype for higher glycinin content, it is conventional to calculate the ratio of 11S:7S (glycinin to β -conglycinin). This ratio was also compared across genotypes. Descriptive statistics for these components are presented in Table 4.1.

With respect to 7S, in developing a line with high S-containing essential amino acids, it is desirable to have the β sub-unit as low as possible since that sub-unit does not contain methionine or cysteine (Krishnan, 2000). The population in this study is promising in that regard, because we identified about 5% of the total lines with less than 6% β sub-unit (Table 4.1).

Table 4.1: Descriptive statistics of soy protein fraction from a population of 101 F6 –derived RIL of N87-984-16 x TN93-99 (average of 2001 and 2003).

Traits	Minimum	Maximum	Mean	Std. Dev.	P1	P2	h ² †
Alpha prime	4.1	7.2	5.6	0.6	6.0	4.9	0.15
Alpha	7.4	11.9	9.5	0.9	10.5	8.1	0.21
Beta	5.6	11.6	8.6	1.1	9.3	6.9	0.38
Conglycinin (7S)	18.8	28.8	23.7	2.2	25.9	19.9	0.30
Acidic	16.8	23.3	19.6	1.2	21.0	16.8	0.29
Basic	12.5	17.9	15.5	1.0	16.4	13.2	0.0
Glycinin (11S)	30.0	39.9	35.1	1.9	37.5	30.0	0.0
Ratio (11S:7S)	1.1	1.8	1.5	0.1	1.5	1.5	0.47
Sum (11S+7S)	49.8	67.0	58.8	3.5	63.3	49.8	0.02

Note: Protein components reported are percent of total protein; P1 = N87-984-16; P2 = TN93-99; †Heritability on an entry mean basis

The α' and α sub-units of 7S do contain traces of methionine and cysteine. Breeders targeting increased S-containing amino acids could select for those components, or develop a selection index, which includes their sum. The sum of α' and α ranged from 11% to 20% with a mean of 15%. This wide range gives us an opportunity to select lines from the upper range. Therefore, selection for a low level of the β sub-unit and high levels of α' α sub units of 7S could help increase the total S-containing amino acids.

Unlike 7S, it is not clear which sub-unit of 11S contains more S-containing amino acids. Both acidic and basic sub-units showed a wide variation among lines. A moderate positive correlation ($r=0.46$; $P<0.01$) was found

between acidic and basic sub-units (Table 4.2). This indicates that both sub-units increase in seeds simultaneously. Thus selection for high levels of both the acidic and the basic sub-units may increase the glycinin level. The level of variation in acidic and basic subunits is reflected in the high level of variation found in glycinin (Table 4.1). The overall distribution of 11S and 7S among the RIL is presented in Figure 4.2 and 4.3 respectively.

Table 4.2: Correlation coefficient between conglycinin, glycinin and their ratio and individual components on the basis of combined least square means from 2001 and 2003.

	Alpha prime	Alpha	Beta	Acidic	Basic	Conglycinin	Glycinin	11S:7S
Alpha	0.63***							
Beta	0.56***	0.42***						
Acidic	0.61***	0.55***	0.33***					
Basic	0.01 ^{ns}	0.14 ^{ns}	0.06 ^{ns}	0.46***				
Conglycinin	0.83***	0.80***	0.85***	0.56***	0.09 ^{ns}			
Glycinin	0.40***	0.43***	0.24**	0.88***	0.82***	0.41***		
11S:7S	-0.64***	-0.59***	-0.76***	-0.04 ^{ns}	0.41***	-0.81***	0.19 ^{ns}	
11S + 7S	0.74***	0.74***	0.67***	0.85***	0.52***	0.86***	0.82***	-0.41***

Note: ns=Non-significant, *=Significant at 0.05 probability, **=Significant at 0.01 probability, and ***=Significant at 0.001 probability.

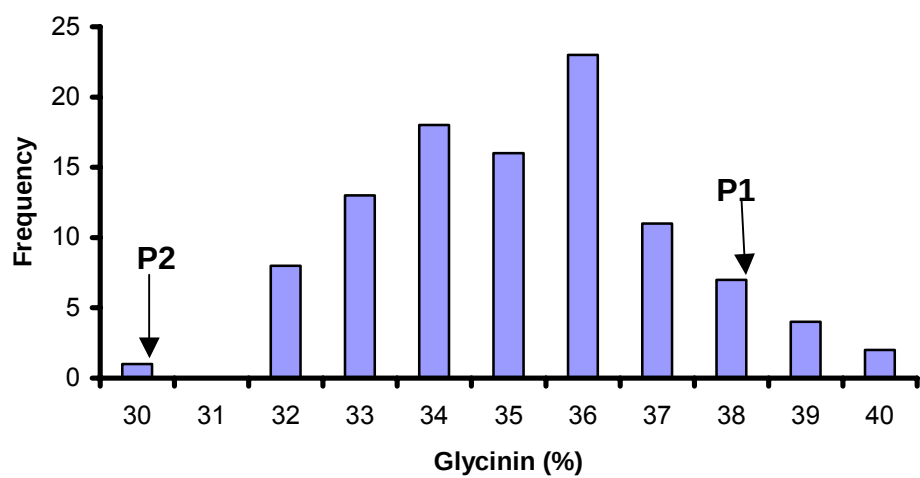


Figure 4.2: Frequency distribution of glycinin in a population of F6 –derived RIL of N87-984-16 x TN93-99.

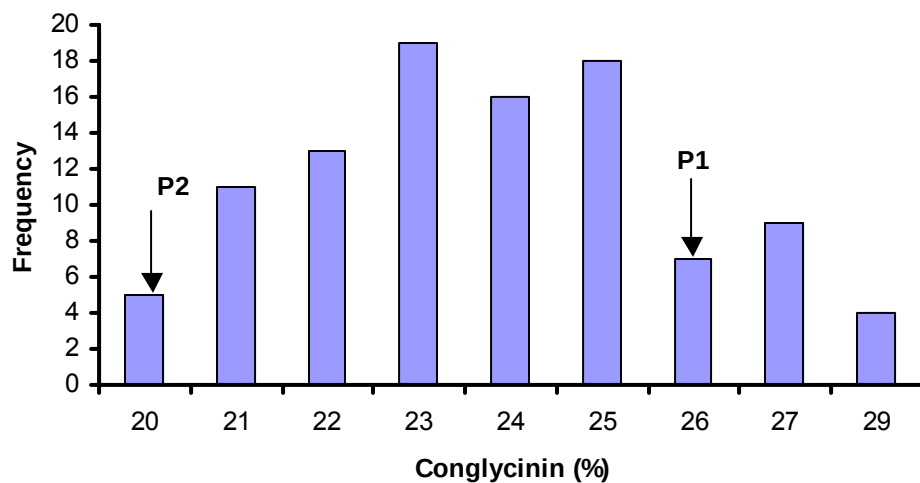


Figure 4.3: Frequency distribution of b-conglycinin in a population of F6 –derived RIL of N87-984-16 x TN93-99.

A high level of variation for 7S and 11S exists in this population. The scatter plot of 11S vs. 7S shows that glycinin content was high in the present population irrespective of β -conglycinin content (Figure 4.4). This indicates that selection for lines with increased 11S could occur among a broad range of 7S levels, thereby providing an opportunity to ensure high protein meal development. Indeed, Prolina appears to contribute to breaking the typical negative correlation between 7S and 11S, because such a correlation was not evident in our population (Figure 4.4).

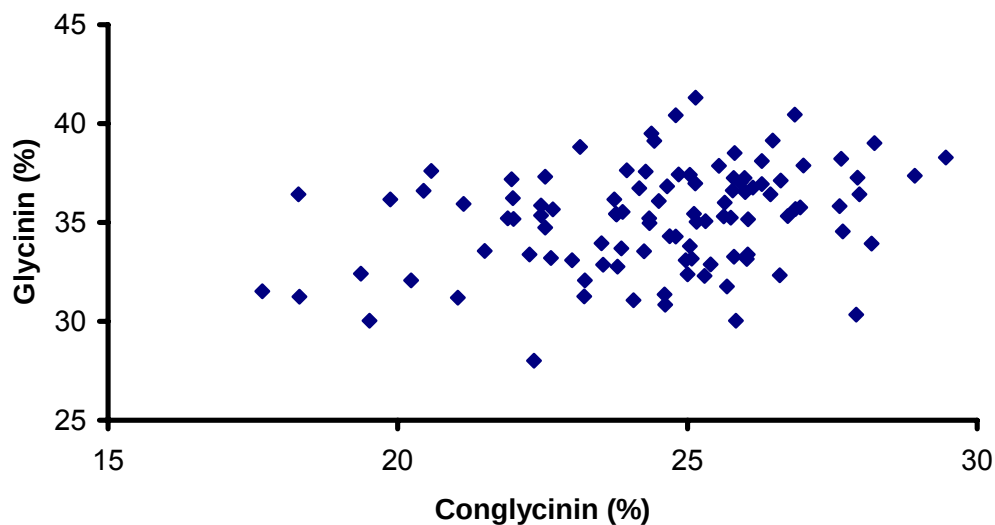


Figure 4.4: Scatter plot of glycinin vs. b-conglycinin in RIL population developed from N87-984-16 x TN93-99. High level of glycinin is present irrespective of b-conglycinin content.

Since 11S contains significantly higher S-containing amino acids, per unit protein than does 7S, an increase in the 11S:7S ratio should lead to improvement in S-containing amino acids (Liu, 1997). Generally, the 11S:7S ratio is less than 1.5 in cultivated soybean but it may be as high as 3.8 in wild *Glycine soja* ancestors of soybean (Kwanyuen et al., 1997; Liu, 1997). The lines with a 11S:7S ratio of 1.8 in the present population (Table 4.1) clearly indicated that we developed RIL having more than typical 11S content. This is likely due to the N87-984-16 (Prolina parent), because Prolina was reported to have higher S-containing peptides than normal cultivars (Wilson, 1996).

Yagasaki et al. (1997) used soybean breeding lines having different glycinin sub-unit composition to determine 11S:7S ratios. The ratio varied from 0.04 to 1.48 among mutant genotypes lacking all three glycinin subgroups. In a similar study So et al. (1999) studied 58 cultivars of soybean according to their groupings (e.g. vegetable, sprout, sauce) and analyzed for different sub-units of storage proteins. They determined the 11S:7S ratio in all cultivars and found that it ranged from 1.4 to 1.7. The cultivars with the highest 11S:7S ratio were in the vegetable type of cultivars. The ratio found in our study was greater than the highest value found by both the groups (So et al., 1999; Yagasaki et al., 1997). This indicates that our population has more potential to develop a cultivar with high S-containing amino acids. Fehr et al. (2003) analyzed 14 soybean

cultivars for various storage proteins for three years at eight locations using different groups of cultivars including three with high protein concentration. They found an 11S:7S ratio of 2.04 only in one high protein cultivar; all others had a lower ratio. This suggests that high protein cultivars *per se* may not have high glycinin, and a need exists to utilize molecular markers (such as the ones we report in this paper) to identify the gene combinations, which may enable improvements in S-containing amino acids.

The QTL for various components and their LOD score are presented in Table 4.3. QTL were detected which showed association between storage protein fractions and its sub-units. Other QTL were detected which are associated with only one component. The QTL near Satt461, located on LG D2, was associated with the α sub-unit of conglycinin in 2001. However, it was not detectable for the α sub-unit in 2003 as in the combined analysis. Therefore, it was environmental specific and not a stable QTL (Brummer et al., 1997). However, the same QTL was associated with the β sub-unit in 2003 and for total conglycinin of storage protein in the combined analysis (Table 4.3).

Since α , α' and β are the sub-units of conglycinin, it is evident that the QTL near Satt461 located on LG D2 is associated with the conglycinin gene family. Another QTL linked to Satt249 located on LG J was detected

Table 4.3: QTL detected for soybean storage protein fractions by composite interval mapping analysis in F6-derived population of N87-984-16 x TN93-99.

Trait	Location	LG†	Marker	QTL Position (cM)	LOD Score	Additive Effect‡	R ² (%)
Alpha	2001	D2	Satt461	93.4	2.1	-0.4	13.4
Alpha prime	2001	K	Satt196	50.3	4.3	3.2	16.7
	2001	K	Satt102	57.3	4.7	4.1	27.0
	2001	L	Satt076	8.0	2.9	-3.3	15.5
	2001	M	Satt540	35.8	3.1	-3.7	19.9
	Combined	A2	Satt177	113.5	2.0	-2.5	14.9
Beta	2003	D2	Satt461	99.4	2.7	-0.6	12.4
	2003	J	Satt249	64.5	2.2	0.6	10.1
	Combined	N	Satt255	0.0	2.3	-0.4	9.1
Conglycinin	2003	J	Satt249	64.5	2.3	1.0	10.3
	Combined	D2	Satt461	99.4	2.1	-0.7	11.6
Acidic	2001	D2	Satt461	71.4	2.8	-0.9	36.6
	2003	D2	Satt461	99.4	2.3	-0.6	11.2
	Combined	D2	Satt461	99.4	3.3	-0.5	14.0
	Combined	K	Satt518	28.0	2.1	0.4	8.1
	Combined	K	Satt196	32.3	2.2	0.4	10.6
	Combined	L	Satt076	10.0	2.5	-0.4	11.7
Basic	2001	A2	Satt133	76.6	2.3	0.6	12.7
	2001	I	Satt292	0.0	3.0	0.6	13.7
	2003	A2	Satt177	117.5	2.6	-0.5	15.9
	2003	L	Satt156	100.5	3.3	0.6	17.6
	Combined	I	Satt292	0.0	2.6	0.4	11.9
	Combined	L	Satt156	104.5	2.1	0.3	10.3
Glycinin	2001	A2	Satt437	65.3	2.6	1.0	12.8
	2001	G	Satt564	14.0	3.7	-1.3	21.8
	2001	G	Satt191	23.2	3.3	-1.2	18.6
	2001	I	Satt292	0.0	2.0	0.8	9.5
	2003	D2	Satt461	99.4	2.3	-1.0	11.2
	2003	L	Satt156	100.5	2.5	1.2	16.0
	Combined	D2	Satt461	87.4	2.3	-0.8	19.5

†LG = Linkage group; ‡½(P1-P2).

during the second year of the study, which was associated with conglycinin and its β sub-unit (Table 4.3). In the combined analysis, QTL near Satt177, located on LG A2, was associated with the α' sub-unit of conglycinin and QTL near Satt255 located on LG N was associated with the β sub-unit. The α , α' and β are the three sub-units of conglycinin (7S) fraction of storage protein. Therefore, it appears that more than one chromosomal region may govern the biosynthesis of conglycinin and its components.

The glycinin of the storage protein is composed of two sub-units acidic and basic. There were four QTL associated with the acidic and two QTL associated with the basic sub-unit of the glycinin fraction of storage protein in the combined data. QTL associated with the acidic sub-unit were near Satt461, Satt518, Satt196 and Satt076, located on LG D2, K, K and L respectively. Since QTL near Satt461 was significant over years and in the combined data, it is a stable QTL associated with the acidic sub-unit of glycinin. The other QTL should be regarded as environmentally sensitive in their control of the acidic sub-unit. QTL associated with the basic sub-unit of glycinin were linked to Satt292 and Satt156 located on LG I and L respectively. Both of these QTL were environmentally stable in determining the concentration of the basic sub-unit of glycinin.

There were six QTL associated with the glycinin fraction of storage protein (Table 4.3). The most noteworthy were near Satt461 (LG D2), which was significant for glycinin in the 2-year combined analysis, and near Satt292 (LG I), which was significant for the basic sub-unit of the glycinin in the 2-year combined analysis. QTL linked to Satt156 appeared to be environmentally sensitive because it was detected only during the second year for both the basic sub-unit and for glycinin. Two environmentally sensitive QTL (near Satt191 and Satt564), both located on LG G were detected only in the first year. The log of odd (LOD) score plot suggests that it may be useful to identify more polymorphic markers between the genomic regions of Satt564 and Satt191 to more finely map the QTL and its position controlling glycinin concentration (Figure 4.5). Detection of QTL with individual components of conglycinin and glycinin but non-significant association with glycinin and conglycinin suggests an interaction of genes governing folding and unfolding of the individual sub-unit of the fractions of the storage protein.

The present experiment identified major QTL for conglycinin, glycinin and their components. These QTL may be useful in conducting marker-assisted selection in soybean breeding programs.

It is believed that 11S is rich in S-containing amino acids hence it is desirable to have greater concentration of 11S. In this experiment, we

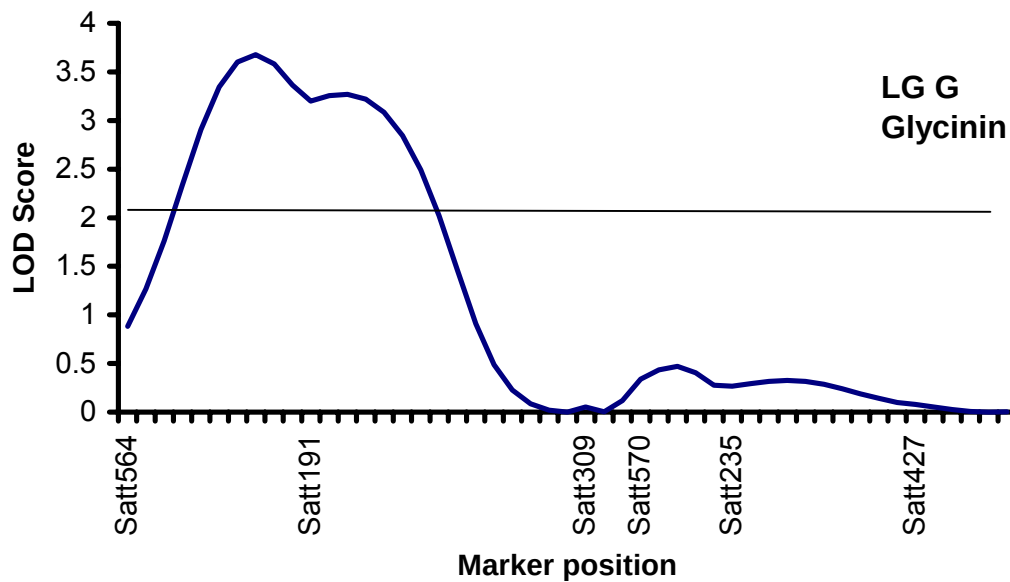


Figure 4.5: LOD score plot for glycinin on LG G. A major QTL appears located between Satt564 and Satt191, and another major QTL appears located downstream of Satt191.

found one common QTL linked to Satt461 for these two fractions. From a MAS point of view, this is not a favorable condition when a negative correlation among protein fractions exists because we want to select only for 11S not for 7S. Therefore, the common QTL near Satt461 cannot be very useful to serve as a unique marker for this purpose. However, there are other QTL near Satt292 and Satt156, associated with 11S but not with 7S. These two QTL can be used as unique QTL in marker assisted selection to improve the nutritional quality of soybean. It should be noted that we attempted to be conservative in reporting only QTL, which were significant based both upon single factor analysis and CIM (Table 4.3).

The environmentally sensitive QTL reported here could be considered stable based upon results from single factor ANOVA. In deed, Stuber et al. (1992) have shown that single factor analysis is as effective as interval mapping, and most soybean protein QTL reported to date have been detected utilizing single factor analysis (www.soybase.org). We anticipate that the QTL reported in this paper will enhance opportunities to improve protein quality in soybean.

A high ratio of 11S:7S may increase the S-containing essential amino acids, hence identification of markers associated with QTL governing 11S and 7S are important to geneticists to identify individuals, which are likely to enable genetic gains for protein quality. Parental allelic contributions for a trait were expressed with respect to P1 (N87-984-16) as an additive effect (Table 4.3). Two markers (Satt156 on LG L and Satt292 on LG I) associated with QTL for glycinin were associated with a positive contribution from P1 whereas Satt461 (LG D2) was contributed from P2 (TN93-99). This indicates that desirable allele frequencies can be increased at those loci to augment genetic gains for glycinin. Therefore the desirable alleles would need to be selected from both the parents and selected individuals could be identified that inherit the best combination of alleles, leading to glycinin levels which exceed that of the higher parent (Figure 4.2). The increase of desirable allele frequencies from both parents in individual RIL is a major goal of modern molecular breeding

programs. For example, Qiu et al. (1999) conducted an experiment and found alleles from both the parents were beneficial for increasing the protein content in a population derived from 'Peking' x 'Essex' soybean.

An interesting aspect of our experiment was that QTL near Satt461 was significantly associated with conglycinin as well as glycinin fractions of seed storage protein. This is not a particularly desirable condition for a marker assisted selection program, because we may want to have reduced 7S and increased 11S content for developing a high S-containing amino acid genotype. If the same marker is associated with both traits, presumably there is pleiotropy or tight genetic linkage. Alternatively, marker Satt292 and Satt156 (located on LG I and L) were found to have a positive contribution for glycinin, and they mapped to a linkage group with no other markers associated with conglycinin. Brummer et al. (1997) found two QTL, located on LG I and L, for protein concentration in soybean but they did not analyze for glycinin. Utilization of our glycinin markers for a marker assisted selection in developing a line with increased 11S:7S ratio may be possible.

Nielsen et al. (1989) identified five genes (*Gy1* through *Gy5*) regulating the production of glycinin in soybean. They grouped the five genes into two groups: three genes in group I and two genes in group II. Nielsen et al. (1989) and Cho et al. (1989) reported that all five genes assort

independently, indicating five different chromosomes. In our experiment, we found molecular markers distributed on three different LGs (D2, I and L) conferring glycinin concentration (Table 4.3). Moreover, Beilinson et al. (2002) found two additional genes (Gy6 and Gy7), which mapped on LG N and L respectively. This indicates that we may not have mapped all QTL in this population, thus further research is warranted. In order to detect more protein quality QTL, development of additional diverse populations would be useful.

Harada et al. (1989) identified three genes, which control the production of β -conglycinin. Furthermore, Nielsen (1996) reported that genes encoding for α and β are linked, whereas α' and α assort independently, indicating that there should be at least two linkage groups containing the genes encoding β -conglycinin. We found two QTL located on two different LG (D2 and J) governing β -conglycinin (Table 4.3).

Increased glycinin content in soybean protein is an important trait for increasing the concentration of the S-containing amino acids. The population developed for this study has been successfully utilized with a molecular breeding approach, resulting in the recent release of soybean germplasm TN04-5363 by the Tennessee Agricultural Experiment Station. This germplasm expresses significantly higher levels of cysteine, and total sulfur-containing amino acid than P1 and P2. It is being made available to

breeders to improve soybean protein quality. The QTL reported in this paper provide new information for soybean geneticists for targeting the development of soybean lines with enhanced protein quality.

References

- Beilinson, V., Z. Chen, R.C. Shoemaker, R.L. Fischer, R.B. Goldberg, and N.C. Nielsen. 2002. Genomic organization of glycinin genes in soybean. *Theor Appl Genet* 104:1132-1140.
- Bradford, M.M. 1976. A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. *Anal Biochem* 72:248-254.
- 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., T.E. Carter, and R.F. Wilson. 1999. Registration of Prolina soybean. *Crop Sci* 39:294-295.
- Cho, T.J., C.S. Davies, and N.C. Nielsen. 1989. Inheritance and organization of glycinin genes in soybean. *Plant Cell* 1:329-337.
- Chua, N.H. 1980. Electrophoretic analysis of chloroplast proteins. *Meths Enzymol* 69:434-446.
- Clarke, E.J., and J. Wiseman. 2000. Developments in plant breeding for improved nutritional quality of soybeans I. Protein and amino acids content. *J Agric Sci* 134:111-124.

- Cregan, P.B., T. Jarvik, A.L. Bush, R.C. Shoemaker, K.G. Lark, A.L. Kahler, N. Kaya, T.T. VanToai, D.G. Lohnes, and J. Chung. 1999. An integrated genetic linkage map of the soybean genome. *Crop Sci* 39:1464-1490.
- Fehr, W.R., J.A. Hoeck, S.L. Johnson, P.A. Murphy, J.D. Nott, G.I. Padilla, and G.A. Welke. 2003. Genotype and environment influence on protein components of soybean. *Crop Sci* 43:511-514.
- Harada, J.J., S.J. Barker, and R.B. Goldberg. 1989. Soybean beta-conglycinin genes are clustered in several DNA regions and are regulated by transcriptional and posttranscriptional processes. *Plant Cell* 1:415-425.
- Hayashi, M., M. Nishioka, K. Kitamura, and K. Harada. 2000. Identification of AFLP markers tightly linked to the gene for deficiency of the 7S globulin in soybean seed and characterization of abnormal phenotypes involved in the mutation. *Breeding Sci* 50:123-129.
- Kitagawa, S., M. Ishimoto, F. Kikuchi, and K. Kitamura. 1991. A characteristic lacking or decreasing remarkably 7S globulin sub-unit induced with γ -ray irradiation in soybean seeds. *Japanese J. Breeding* 41:460-461.
- Kitamura, K. 1995. Genetic improvement of nutrition and food processing quality in soybean. *Japanese Agric Res Quart* 29:1-8.

- Kitamura, K., and N. Kaizuma. 1981. Mutant strains with low level of subunits of 7S globulin in soybean (*Glycine max* Merr.) seed. Japanese J Breeding 31:353-359.
- Krishnan, H.B. 2000. Biochemistry and molecular biology of soybean seed storage proteins. J New Seeds 2:1-25.
- Kwanyuen, P., and R.F. Wilson. 2000. Optimization of coomassie staining for quantitative densitometry of soybean storage proteins in gradient gel electrophoresis. J Am Oil Chem Soc 77:1251-1254.
- Kwanyuen, P., V.R. Pantalone, J.W. Burton, and R.F. Wilson. 1997. A new approach to genetic alteration of soybean protein composition and quality. J Am Oil Chem Society 74:983-987.
- Liu, K. 1997. Soybeans: chemistry, technology and utilization Chapman and Hall.
- Luck, P., T. Lanier, C. Daubert, R.F. Wilson, and P. Kwanyuen. 2001. Functionality and viscoelastic behavior of Prolina soybean isolate, p. 197-202, *In* R. F. Wilson, ed. Oilseed processing and utilization. AOCS Press, Champaign, IL.
- Mandal, S., and R. Mandal. 2000. Seed storage proteins and approaches for improvement of their nutritional quality by genetic engineering. Curr Sci 79:576-589.
- Nielsen, N. 1996. Soybean seed composition, p. 127-163, *In* D. Verma and R. Shoemaker, eds. Soybean: Genetics, molecular biology and biotechnology. CAB International, Wallingford.

- Nielsen, N.C., C.D. Dickinson, T.J. Cho, V.H. Thanh, B.J. Scallan, R.L. Fischer, T.L. Sims, G.N. Drews, and R.B. Goldberg. 1989. Characterization of the glycinin gene family in soybean. *Plant Cell* 1:313-328.
- Ogawa, T., E. Tayama, K. Kitamura, and N. Kaizama. 1989. Genetic improvement of seed storage proteins using three variant alleles of 7S globulin sub-unit in soybean (*Glycine max* L.). *Japanese J Breeding* 39:137-147.
- Pantalone, V.R., F.L. Allen, and D. Landau-Ellis. 2003. Registration of 'TN93-99' Soybean germplasm. *Crop Sci* 43:1137.
- Qiu, B.X., P.R. Arelli, and D.A. Sleper. 1999. RFLP markers associated with soybean cyst nematode resistance and seed composition in a 'Peking' x 'Essex' population. *Theor Appl Genet* 98:356-364.
- SAS Institute Inc. 2000. SAS guide for personal computers. 8th Edition ed. SAS Institute, Cary, NC.
- So, E., Y. Chae, Y. Kim, and Y. MH. 1999. Variation of 7S and 11S seed protein concentrations in different food types of soybean seed. *Korean J Crop Sci* 44:350-354.
- Stuber, C., S. Lincoln, D. Wolff, T. Helentjaris, and E. Lander. 1992. Identification of genetic factors contributing to heterosis in a hybrid from two elite maize inbred lines using molecular markers. *Genetics* 132:823-839.

- Takahashi, K., Y. Mizumo, S. Yumoto, K. Kitamura, and S. Nakamura. 1996. Inheritance of the α -sub-unit deficiency of β -conglycinin in soybean (*Glycine max* L) line induced by γ -ray irradiation. *Breeding Sci* 46:251-255.
- Tsukada, Y., K. Kitamura, K. Harada, and N. Kaizuma. 1986. Genetic analysis of two major sub-units (β -conglycinin and glycinin) in soybean seeds. *Japanese J Soybean* 36:390-400.
- Wilson, R.F. 1996. Current research on oilseeds Institute of Food Technologists 1996 Symposium: Identify Preserved Oils. Institute of Food Technologists, New Orleans, LA.
- Yagasaki, K., T. Takagi, M. Sakai, and K. Kitamura. 1997. Biochemical characterization of soybean protein consisting of different subunits of glycinin. *J Agric Food Chem* 45:656-660.
- Yaklich, R.W. 2001. β -Conglycinin and glycinin in high-protein soybean seeds. *J Agric Food Chem* 49:729-735.

PART 5

Quantitative Trait Loci Controlling Sulfur Containing Amino Acids, Methionine and Cysteine, in Soybean Seeds

This chapter is a lightly revised version of paper by the same name that has been submitted to the *Theoretical and Applied Genetics* in 2004 by D.R. Panthee, V.R. Pantalone, C.E. Sams, A.M. Saxton, D.R. West, J.H. Orf and A.S. Killam

My use of “our” in this chapter refers to my co-authors and myself. My primary contribution to this paper include (1) organization, development and management of all aspects of the project, (2) collecting all molecular data using polymorphic molecular markers, (3) collecting all the field data at Knoxville and Holston environments, (4) analyzing for amino acids composition, (5) analyzing all data, and (6) all of the writing except for some minor revisions.

Abstract

Soybean [*Glycine max* (L.) Merr.] is the single largest source of protein in animal feed. However, a major limitation of soy proteins is their deficiency of sulfur-containing amino acids, methionine (Met) and cysteine (Cys). The objectives of this study were to determine genotypic differences for Met and Cys and to identify genomic regions governing Met and Cys composition in soybean seed. To achieve these objectives, 101 F₆-derived recombinant inbred lines (RIL) from a population developed from a cross of N87-984-16 x TN93-99 were used. Ground soybean seed samples were analyzed for Met and Cys composition using a near infrared

spectroscopy (NIRS) instrument. Data were analyzed using SAS software and QTL Cartographer. RILs differed ($P < 0.05$) in Met and Cys concentrations, with a range of 5.1 to 7.3 (g kg^{-1} seed dry weight) for Cys and 4.4 to 8.8 (g kg^{-1} seed dry weight) for Met. Heritability estimates on entry mean basis were 0.14 and 0.57 for Cys and Met, respectively. A total of 94 polymorphic simple sequence repeat (SSR) molecular genetic markers were screened in the RIL. Single factor ANOVA was used to identify candidate QTL, which were confirmed by composite interval mapping with QTL Cartographer. Nine QTL linked to molecular markers Satt002, Satt185, Satt201, Satt235, Satt252, Satt268, Satt427, Satt436 and Satt590 distributed on six molecular linkage groups were associated with Cys and four QTL linked to molecular markers Satt002, Satt252, Satt564 and Satt590 distributed on four molecular linkage groups were associated with Met concentration in soybean seed. QTL associated with Met and Cys in soybean seed will provide important information to breeders targeting improvements in the nutritional quality of soybean.

.Key words: Amino acid, Cys, genomic regions, *Glycine max*, heritability, Met, QTL mapping, soybean.

Introduction

Soybean [*Glycine max* (L.) Merr.] is a major source of protein for humans and livestock throughout the world. It is estimated that more than 70% of the world's total protein meal comes from soybean (Soytech Inc., 2004). A

major function of proteins in nutrition is to supply adequate amounts of required amino acids (Friedman and Brandon, 2001). Although soybean protein is relatively rich in most of the essential amino acids, the concentration of the sulfur containing amino acids methionine (Met) and cysteine (Cys) is approximately half that of egg protein, which is the standard nutritional reference protein (George and de Lumen, 1991) identified by the Food and Agriculture Organization (FAO) of United Nations. Because of this deficiency, either synthetic or natural supplementary ingredients are utilized to fulfill the requirement of Met in soy based animal feed. However, Met supplementation has possible problems such as leaching during processing and bacterial degradation leading to formation of undesirable volatile sulfides (George and de Lumen, 1991).

Met occupies a central position in cellular metabolism in which the processes of protein synthesis, methyl group transfer, polyamine and ethylene syntheses are interlocked (Ravanel et al., 1998). Among these pathways, the synthesis of proteins is the only pathway consuming the entire Met molecule. Since Met is involved in all these physiological and biochemical processes, ample supply of Met in animal feed is needed. The FAO recommended a combined requirement of Cys and Met as 3.5% of total protein (George and de Lumen, 1991). In regular soybean cultivars, average Met content is 0.65g/100 g of dry matter (Friedman and

Brandon, 2001). Efforts are needed to develop soybeans with enhanced Met to overcome these problems.

The amount of supplemental Met required to balance a deficient ration for animals is about 4 to 10 g per day per animal. Sulfur containing amino acids are extremely important in wool producing animals to produce high quality wool (Liu and Masters, 2003). Natural sources of Met (such as corn gluten meal or animal proteins) supply Met at a cost of about US\$ 0.03 g⁻¹. Synthetic rumen-protected Met products typically supply Met at about US\$ 0.02 g⁻¹. It is estimated that the commercial value of a 10% improvement in the present level of Met would be about as US\$3.5 t⁻¹ (Clarke and Wiseman, 2000). Since most cereals are rich in S-containing amino acids, the deficiency of S-containing amino acids in soybean is overcome by complementation with cereals and/or by synthetic supplements in the poultry and swine industry. The poultry and swine industry would derive considerable savings if an improved cultivar of soybean supplementation of cereal-soybean meal based animal feed with exogenous Met. Some people, whose source of food is root crops e.g. cassava (*Manihot esculenta* Crantz), suffer from deficiency of S-containing amino acids. If a genetic solution could be provided in soybean, it would also help to solve this deficiency problem.

Efforts made in the past to address low Met include conventional plant breeding, biotechnological approaches, and nutrient amendment. One of the successful efforts was development of mutants with increased concentration of Met and Cys by the use of 0.5% Ethylmethanesulfonate (EMS) (Imsande, 2001). He was able to develop a mutant line with both Met and Cys concentrations of 1.85% of total protein. Another research group was able to increase the Met concentration in maize inbreds by 17% over their recurrent parent via backcrossing (Olsen et al., 2003).

Muntz et al. (1998) have discussed in detail the approaches adopted to increase concentration of essential amino acids including Met and Cys by means of biotechnology. The most promising approach adopted was the introduction of foreign gene(s) encoding for proteins with extremely high Met level. Examples of such approaches are the gene encoding for 2S albumin of Brazil nut (*Bertholettia excelsa*) (Altenbach et al., 1987) and sunflower (*Helianthus annuus*) (Kortt and Caldwell, 1990). The feasibility of this strategy was shown by transforming tobacco (Altenbach et al., 1989). The transformation approach was also successfully used in soybean (Townsend and Thomas, 1994). The Met concentration in transgenic soybean was nearly double that of the wild type. However, the Brazil nut allergen was also transferred making the transgenic beans unusable (Nordlee et al., 1996). Despite all of these efforts, not a single commercial

cultivar of soybean with FAO standard total sulfur containing amino acids has been developed to date.

A nutrient amendment approach to improve sulfur containing amino acids included regulation of N and S in soil (Imsande and Schmidt, 1998; Sexton et al., 1998). They assessed the seed quality of soybean with respect to 7S and 11S fractions of seed storage protein because 11S contains more Met concentration than 7S.

For a breeding program to focus on increased levels of Met and Cys, hundreds of lines in a population must be assessed for amino acid composition. Various methods have been employed for this purpose. A few are use of bacteria (e.g. *Pediococcus cerevisiae*, which is auxotrophic for Met) (Wright and Orman, 1995), High Pressure Liquid Chromatography (HPLC) (Zarkadas et al., 1999; Zarkadas et al., 1997) and near-infrared reflectance spectroscopy (NIRS) (Bezant et al., 1997; Fontaine et al., 2001; Pazdernik et al., 1997). In this paper, we report the quantitative trait loci (QTL) for Met and Cys of soybean protein analyzed using NIRS.

Materials and Methods

Plant material

A total of 101 F_6 -derived recombinant inbred lines (RIL) were developed from a cross of N87-984-16 (P1) x TN93-99 (P2). The N87-984-16 line is one of two F_8 -derived near isogenic sister lines, whose blend constitutes the high protein commercial cultivar 'Prolina' (Burton et al., 1999). The line TN93-99 is a high yielding and well adapted University of Tennessee breeding line and is currently registered as germplasm GP-280 (Pantalone et al., 2003). Crossing was done during summer of 1998. F_1 seeds were harvested in October 1998 and F_2 single plants were grown in Costa Rica during winter 1998. Generations were advanced via single seed descent until F_5 in Costa Rica and $F_{5:6}$ seeds were obtained in May 2000. About 300 $F_{5:6}$ single plants were grown at the Knoxville Experiment Station in Knoxville, TN as a recombinant inbred line (RIL) population and 101 random plants of Tennessee adapted maturity were selected as a source population for this study. All 101 lines and parents were planted in 6m length two-row plots with three replications in a randomized complete block design (RCBD) at the Knoxville Plant Science Farm (KPSF) in 2001. All the RIL, the two parents, as well as checks (Hutcheson, 5002T and 5601T) were planted in a RCBD with three replications at three locations: the Knoxville and Holston unit of the Knoxville Experiment Station of the University of Tennessee, and the Ames Experiment Station, Grand

Junction, TN in May 2002 and 2003. Each line was planted in a four row plot of 6 m length with a spacing of 75 cm between rows.

Sample preparation and amino acid analysis

Approximately 20 g of soybean seeds were ground in a water-cooled Knifetec 1095 Sample Mill (FOSS Tecator, S-26321, Hogana, Sweden) for 20 seconds. This setting produced soybean flour with a uniform particle size. The NIRS instrument (NIRS 6500, FOSS North America) was warmed up for 2 h after turning on the lamp and auto diagnostics were run. diagnostics tests ensured the instrument passed three different tests for instrument response, wavelength accuracy, and NIRS repeatability. A room dehumidifier was used throughout the analysis, setting the humidity to 40%; room temperature was approximately 20°C. The ground soybean samples were scanned using Winisi II 1.5 software. The instrument was left on for the entire period of analysis, and diagnostics were performed every day until scanning of all samples was completed. The scanning produced the predicted levels of Met and Cys as concentration of total seed on a dry weight basis. We also calculated total sulfur containing amino acids (Met + Cys) and their ratio (Met: Cys).

DNA extraction and polymerase chain reaction

DNA was extracted from the RIL and parental lines utilizing Qiagen Plant Easy DNA Extraction Kit (Qiagen Huden, Germany). Polymerase chain

reaction (PCR) consisted of 7.4 μL of ddH₂O, 1 μL of 10X PCR Buffer, 1 μL of 2 mM dNTPs mixture (Pharmacia, Piscataway, NJ), 0.5 μL of 20 μM forward and reverse primer, 0.1 μL of 5 units/ μL KlenTaq (Ab Peptides Inc., St. Louis, MO) and 2 μL of 20 ng/ μL template DNA. The PCR was performed in a 96-well MBS Hybaid thermocycler (Hybaid, Franklin, MA). PCR conditions were a) 94°C for five minutes followed by 35 cycles at 94°C for denaturation for 25 s, b) 47°C for annealing for 30 s, c) 72°C for 25 s for extension, and one last cycle at 72°C for final extension for five minutes. Parents were screened for a total of 585 (ATT)_n simple sequence repeat (SSR) genetic markers (Cregan et al., 1999), of which 138 were found polymorphic. The sequence information for the markers is publicly available from the Website (www.soybase.org), which was verified on December 16, 2004. A total of 94 scorable and polymorphic SSR markers were used in QTL analyses.

DNA gel electrophoresis

A 6% non-denaturing polyacrylamide gel electrophoresis (PAGE) used to separate the PCR product consisted of 6% bis-acrylamide, 0.5% TBE buffer, 0.07% APS and 0.035% TEMED, which was measured as 28.5 mL of 40% bis-acrylamide, 160.2 mL of 0.593X TBE buffer, 1.33 mL of 10% APS and 66.5 μL of TEMED. Two μL of loading buffer (6X) were added to PCR products and a 10 μL sample was loaded on the gel. The running buffer was 0.5X TBE. The gel was run at constant 300 V for 3 h. A fan was

used to keep the glass plates cool during gel running. Ethidium bromide (50 µl, 10 mg/mL) was added to the running buffer in order to visualize the bands under exposure to UV light. Bands were scored using 1 to represent P₁ (N87-984-16), 2 to represent heterozygote, and 3 to represent P₂ (TN93-99) alleles for each primer locus.

Data analysis

Phenotypic data on Met and Cys were analyzed using the MIXED procedure of SAS software to determine genotypic differences among the RIL. Location and replication were considered as two blocking factors in the model. Heritability of the trait in the population was estimated on entry mean basis (Nyquist, 1991) as follows:

$$h^2 = \frac{\sigma_g^2}{\sigma_g^2 + \left(\frac{\sigma_{ge}^2}{e} \right) + \left(\frac{\sigma^2}{re} \right)}$$

where h^2 represents the heritability, σ_g^2 is genotypic variance, σ_{ge}^2 is genotype x environment variance, σ^2 is error variance, r is number of replications and e is the number of environments.

REML was used to estimate variance components for calculating heritability estimates. Phenotypic correlation was determined using the CORR procedure and genetic correlation was determined using the following formula (Falconer and Mackay, 1996; Kearsy and Pooni, 1996):

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

where, r_G represents genetic correlation, x represents one trait, y represents a second trait and σ^2 is genetic variance. Cross products (Cov) were generated using the MANOVA option of the PROC GLM procedure. Associations between molecular data and least squares means of NIRS predicted Met and Cys were detected by analyzing the data using PROC GLM (SAS Institute Inc, 2002). heterozygotes were excluded from the analysis because the RIL population was F_6 -derived, and our interest was to detect heritable alleles from pure lines. QTL position and distance were estimated using Mapmaker/Exp 3.0 (Lincoln et al., 1993) and the information from Mapmaker was then used in QTL Cartographer (Wang et al., 2003) to confirm the QTL by composite interval mapping (CIM). We used the standard model *Zmapqtl* 6 in the CIM procedure with a 10 cM window size and a 2 cM walking speed. We performed 1000 permutations on Cys, Met and Met+Cys in each environment and in the combined data from six environments to establish the empirical LOD thresholds at the 5% probability level (Churchill and Doerge, 1994). However, QTL with LOD $\geq$ 2.0 are also reported as in other papers (Concibido et al., 2003; Njiti et al., 2002; Rector et al., 1998) because these possible QTL may guide other researchers in future genetic studies.

Results and Discussion

Methionine and cysteine phenotype

A significant difference ($p < 0.01$) was found among genotypes in the soybean population for Met and Cys. Descriptive statistics (g kg^{-1} seed dry weight) for Met, Cys and a combination of Met + Cys are presented in Table 5.1. The frequency distributions for Met, Cys and Met+Cys are presented in Figures 5.1 through 5.3, respectively. RIL means ranged from 5.1 to 7.3, and from 4.4 to 8.8 (g kg^{-1} seed dry weight) for Cys and Met respectively. The combination of total S-containing amino acids (Met + Cys) ranged from 9.7 to 15.9 (g kg^{-1} seed dry weight) among the RIL. The upper end of these ranges represents a high level of S-containing amino acid concentration in soybean. The average concentrations of Met and Cys in soybean reported by Friedman and Brandon (2001) are 0.65% each of dry matter, which is 6.5 g kg^{-1} seed dry weight making total S-containing amino acids 13.0 g kg^{-1} of dry matter. The combination of Met + Cys was 1.28% of dry matter (12.8 g kg^{-1} seed dry weight) in Fontaine et al. (2001). Kwanyuen et al. (1997) analyzed the amino acid composition in selected *Glycine soja*, a wild relative of cultivated soybean germplasms by HPLC method and found that the concentration of Met and Cys ranged from 0.7 to 1.1% (7 to 11 g kg^{-1} dry weight) and 0.4 to 0.8% (4 to 8 g kg^{-1} seed dry weight) respectively. On the basis of these past observations, the average Met and Cys levels found in our population were quite high.

Table 5.1: Descriptive statistics (seed dry weight basis) for cysteine (Cys), methionine (Met), total sulfur containing amino acids (Met+Cys) and total seed protein concentration averaged over six environments in a soybean RIL population derived from N87-984-16 x TN93-99.

Trait	Minimum	Maximum	Mean	P1	P2	LSD _{0.05}	h ² †
Cysteine (g kg ⁻¹)	5.1	7.3	6.2	7.0	6.0	0.3	0.14
Methionine (g kg ⁻¹)	4.4	8.8	6.4	6.8	6.8	0.7	0.57
Cys + Met (g kg ⁻¹)	9.7	15.9	12.6	13.8	12.8	0.7	0.42
Protein (g kg ⁻¹)	399.7	434.5	416.1	431.0	401.0	3.3	0.66

P1 = N87-984-16; P2 = TN93-99; LSD = Least significant difference at the 0.05 probability level; †Heritability on an entry mean basis

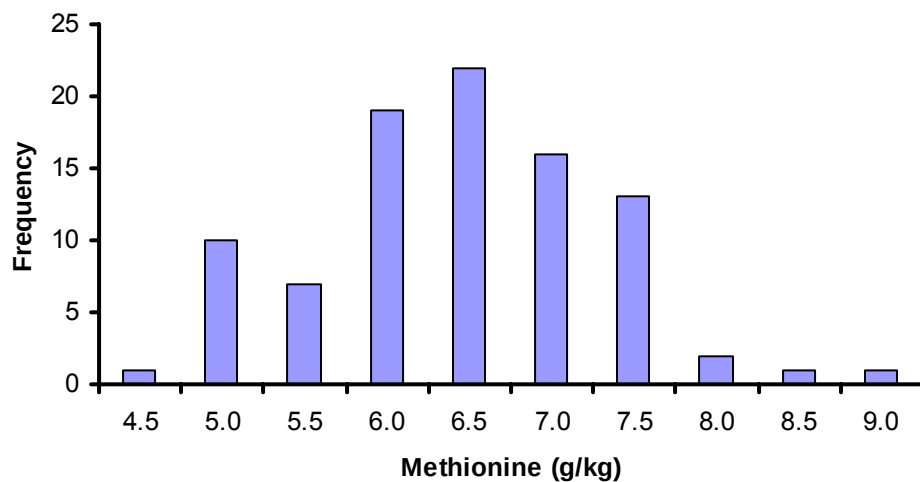


Figure 5.1: Distribution of Met (g kg⁻¹ seed dry matter) in an F6-derived RIL population of soybean developed from a cross of N87-984-16 x TN93-99.

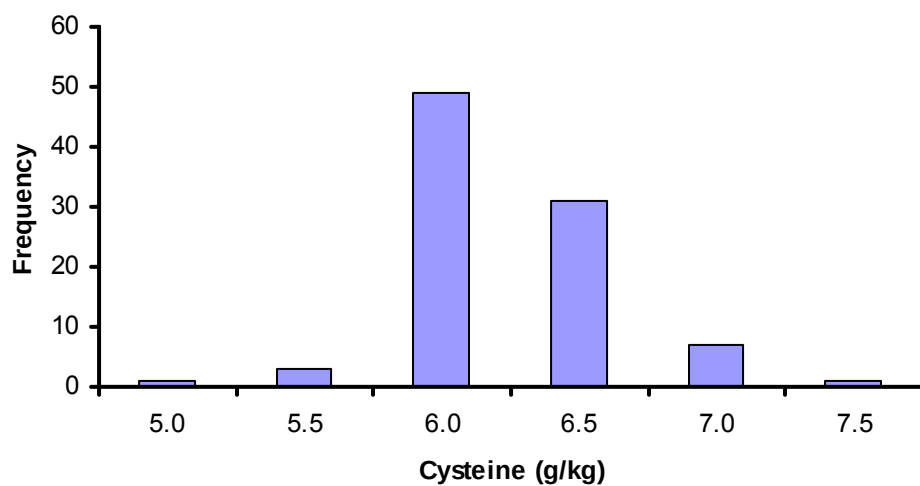


Figure 5.2: Distribution of Cys (g kg⁻¹ seed dry matter) in an F6-derived RIL population of soybean developed from a cross of N87-984-16 x TN93-99.

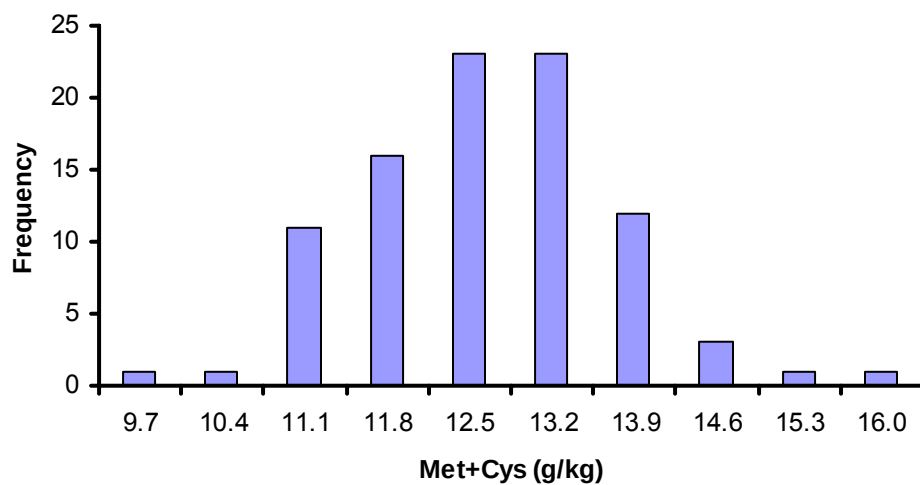


Figure 5.3: Distribution of Met+Cys (g kg⁻¹ seed dry matter) in an F6- derived RIL population of soybean developed from a cross of N87-984-16 x TN93-99.

The available variation and observed phenotypes for Met and Cys should be interpreted with respect to heritability. A moderately high heritability was found for Met (0.57) in data combined from six environments, whereas heritability was low (0.14) for Cys. Quantitative traits like Met and Cys are influenced by the growing environments to a great extent. Whitehead et al. (1989) concluded from a similar type of study that a high level of heritability observed for amino acids indicated that phenotypes observed are likely to have similar observations in future generations. selection from among high performing lines is more likely to achieve significant genetic gains when heritabilities are high. Unlike qualitative traits, it is difficult to make significant response to selection in quantitative traits if the heritability is low.

According to the FAO standard, the combination of Met + Cys should be $\geq 3.5\%$ of protein (approximately 14.0 g kg^{-1} seed dry weight) to make the soybean protein equivalent to egg protein, a reference protein (Clarke and Wiseman, 2000). Typically, soy protein contains only 0.56% of the S-containing amino acids relative to egg protein (Wilson, 1987). In the present study, the combined average Met+Cys was 12.6 g kg^{-1} , ranging from 9.7 to 15.9 g kg^{-1} seed dry weight, greater than the FAO standard. Generally, both Met and Cys were higher in this population, but Cys was much higher than in other populations. One of the parents (N87-984-16) constitutes the high protein commercial cultivar 'Prolina', which contains

not only high protein but also exhibits superior gelling properties of its storage protein, indicating that it may contain more S-containing amino acids (Kwanyuen et al., 1998; Luck et al., 2001). Our choice of N87-984-16 as one of the parents provided us with an opportunity to generate lines with higher S-containing amino acids. Recently, TN04-5363 was registered as an improved S-containing amino acid germplasm from this population (Panthee and Pantalone 200x).

We used NIRS to estimate the Met and Cys concentration in this soybean population. Pazdernik et al. (1997) analyzed soybean samples for amino acids by HPLC and NIRS instruments to validate their prediction equation. They found very close agreement by both methods, showing the validity of the NIRS method to predict the amino acid concentration. Therefore, the NIRS predicted value in the present population was expected to emulate HPLC measurements. In our population, a negative phenotypic correlation between protein and Cys ($r=-0.48$) and no correlation between protein and Met ($r=-0.13^{ns}$) (Table 5.2) indicated that high protein genotypes *per se* may not have high S-containing amino acids, which is in agreement with past findings (Whitehead et al., 1989). Therefore it is necessary to assess the lines for S-containing amino acids in a breeding program aiming to improve protein quality. The NIRS is fast, easy and accurate (Fontaine, 2003) and serves the purpose well for applied breeding research.

Table 5.2: Genetic and phenotypic correlation coefficient among Cys, Met and other traits in an F6-derived RIL population of soybean from N87-984-16 x TN93-99.

Trait	Cysteine		Methionine		Summc	
	Genetic	Phenotypic	Genetic	Phenotypic	Genetic	Phenotypic
Maturity	-0.90***	-0.64***	-0.84***	-0.48***	-0.94***	-0.68***
Height	-0.44***	-0.300**	-0.67***	-0.35**	-0.54***	-0.380**
Seed wt	-0.41***	-0.300**	-0.43***	-0.250**	-0.44***	-0.330**
Yield	-0.030 ^{ns}	0.006 ^{ns}	0.48***	0.170 ^{ns}	0.140 ^{ns}	0.080 ^{ns}
Oil	0.110 ^{ns}	0.020 ^{ns}	0.100 ^{ns}	0.100 ^{ns}	0.110 ^{ns}	0.060 ^{ns}
Protein	-0.64***	-0.48***	-0.250**	-0.130 ^{ns}	-0.55***	-0.40***
Methionine	0.73***	0.44***	-	-	-	-
Summc	0.97***	0.91***	0.86***	0.76***	-	-
Ratiomc	-1.00***	-0.93***	-0.64***	-0.150 ^{ns}	-0.94***	-0.73***

*, ** and *** = Significant at 0.05, 0.01 and 0.001 probability levels respectively; ns = Non-significant; Summc = Met + Cys; Ratiomc = Met:Cys

The population was grouped into two subsets of relative plant maturity for ease of field production and to minimize measurement errors. The subsets, called maturity A and B were approximately different by two weeks. It is interesting to note that there are six clear clusters in the scatter diagram of total S-containing amino acid (Met + Cys) (Figure 5.4). Evaluation of individual lines and Met and Cys concentration revealed that the earlier maturing genotypes (maturity A) produced higher S-containing amino acids in all three locations than maturity B. There was a negative phenotypic correlation between maturity and total S-containing amino acid content ($r=-0.68$) in this population. This is the first report of such an association in soybean.

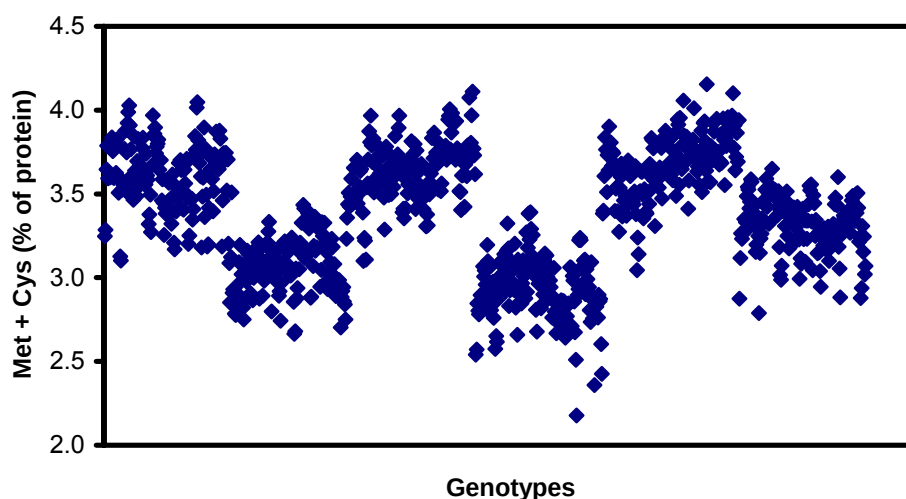


Figure 5.4: Scatter diagram of total S-containing amino acid (% of protein) in a soybean RIL population derived from N87-984-16 x TN93-99.

QTL Analysis

CIM analysis detected nine QTL linked to molecular markers Satt002, Satt185, Satt201, Satt235, Satt252, Satt268, Satt427, Satt436 and Satt590 significantly associated with Cys in this population. These QTL were distributed on six molecular linkage groups D1a, D2, E, F, G and M respectively (Table 5.3). Out of nine, four QTL near Satt235, Satt252, Satt427 and Satt436 showed association with Cys in multiple environments indicating that these may be stable QTL. QTL linked to Satt185 and Satt268 on LG E are within 0.5 cM distance of each other and had similar genetic effects, indicating that either marker is associated with the same QTL in that genomic region.

Table 5.3: QTL detected by composite interval mapping for cysteine, methionine and Cys + Met in F6-derived RIL population of soybean from N87-984-16 x TN93-99.

Trait	Environment	LG	Marker	Position (cM)	LOD Score	Additive Effect‡	R ² (%)
Cysteine	Holston 2002	D1a	Satt436	70.69	2.64 [†]	-0.02	9.5
	Holston	D1a	Satt436	70.69	2.46	-0.02	8.5
	Holston 2002	D2	Satt002	47.73	2.55 [†]	-0.02	9.6
	Holston 2003	E	Satt185	44.76	2.99 [†]	-2.42	10.8
	Holston 2003	E	Satt268	44.27	2.51	-2.29	9.6
	KPSF	F	Satt252	16.08	2.20	1.52	10.1
	Combined 2003	F	Satt252	16.08	2.66 [†]	1.49	12.4
	Combined	F	Satt252	16.08	2.22	0.01	11.0
	Holston 2003	G	Satt235	21.89	2.62 [†]	2.25	9.4
	Combined 2003	G	Satt235	21.89	2.63 [†]	1.46	13.0
	Combined	G	Satt235	21.89	2.78 [†]	0.01	12.6
	Ames	G	Satt427	51.69	2.29	0.01	12.1
	KPSF 2003	G	Satt427	51.69	3.50 [†]	2.14	13.8
	KPSF 2003	M	Satt201	13.56	2.18	1.66	8.1
	KPSF 2003	M	Satt590	7.84	2.32	-1.68	8.1
Methionine	Holston 2002	D2	Satt002	47.73	3.48 [†]	-0.04	12.6
	KPSF	F	Satt252	16.08	2.84 [†]	0.47	15.2
	Holston 2003	G	Satt564	57.32	3.38 [†]	-0.75	24.5
	KPSF 2003	G	Satt564	57.32	2.37	-0.63	14.9
	KPSF	G	Satt564	57.32	2.34	-0.50	18.0
	Combined 2003	G	Satt564	57.32	2.59 [†]	-0.54	18.8
	Ames	M	Satt590	7.84	2.84 [†]	-0.04	20.5
	Holston 2003	M	Satt590	7.84	2.62 [†]	-0.61	16.2
	Holston	M	Satt590	7.84	3.28 [†]	-0.05	19.8
	Combined 2003	M	Satt590	7.84	2.47	-0.55	19.2
Met+Cys	Combined	M	Satt590	7.84	2.42	-0.04	22.9
	Holston 2002	D2	Satt002	47.73	3.23 [†]	-0.06	11.7
	Holston	D2	Satt002	47.73	2.10	-0.04	7.6
	Holston 2003	D2	Satt461	80.19	5.02 [†]	-1.20	49.7
	Holston 2003	E	Satt185	44.76	3.39 [†]	-0.59	11.9
	Holston 2003	E	Satt268	44.27	2.02	-0.46	7.4
	KPSF 2002	F	Satt252	16.08	2.82 [†]	0.57	13.9
	KPSF	F	Satt252	16.08	2.88 [†]	0.61	16.1
	Combined 2002	F	Satt252	16.08	2.04	0.04	10.7
	Combined	F	Satt252	16.08	2.45	0.04	12.3
	Holston 2003	G	Satt564	57.32	2.56	-0.63	13.1
	Ames	M	Satt590	7.84	2.43	-0.05	17.6

LG = Linkage group; LOD = Log of odds ratio; †Significant QTL at empirical threshold levels determined by 1000 permutations; ‡ ½(P1-P2).

Although Satt201 and Satt590 on LG M are within 6 cM distance, their genetic effect was in opposite direction, indicating that they may be associated with different QTL. Satt002 was detected only in one environment indicating that it may be associated with an environmentally sensitive QTL (Brummer et al., 1997). Therefore, it appears that eight QTL were associated with Cys concentration in the present study.

Out of eight QTL, four of the loci linked to Satt201, Satt235, Satt252 and Satt427 received increasing alleles for Cys through P1 whereas the remaining four loci received positive contributions through P2. This indicates that Cys contributing alleles were not concentrated in one parent, a common phenomenon in quantitative traits. On the basis of the magnitude of additive genetic effects QTL near Satt185, Satt235, Satt268 and Satt427 were important loci. Phenotypic variation explained by individual QTL ranged from 8.1 to 13.8% (Table 5.3).

Among the six linkage groups detected as governing Cys concentrations, D1a, E, F, G and M have already been reported to have QTL associated with protein (www.soybase.org), although no QTL for Cys have been reported. Thus, our information is novel and important for future soybean Cys genetic research.

In our population, genetic gain for Cys was contributed by both parents but the magnitude of contributions were higher for P1. This is as expected because P1 was thought to be rich in S-containing amino acids (Kwanyuen et al., 1998; Luck et al., 2001). Moreover, N87-984-16 has some unusual biochemistry in its β -conglycinin (7S) fraction of protein in the sense that it had almost five times more Cys concentration (Kwanyuen et al., 1998), which is presumably because of sulfur-bonding. That may explain why more Cys was contributed via P1 alleles.

There were four QTL linked to molecular markers Satt002, Satt252, Satt564 and Satt590 significantly associated with Met concentration in this population, which were distributed on molecular linkage groups D2, F, G and M, respectively (Table 5.3 and Figure 5.5). Among these QTL, those near Satt564 and Satt590 were associated with Met concentration in multiple environments indicating that they may be the stable QTL. QTL near Satt002 and Satt252 were detected only in one environment indicating that these may be environmentally sensitive (Brummer et al., 1997). Interestingly, unlike in Cys, three of the four loci (near Satt002, Satt564 and Satt590) received increasing alleles through P2 (Table 5.3). Phenotypic variation explained by an individual QTL ranged from 12.6 to 24.5% indicating that all the QTL detected were major QTL (Falconer and Mackay, 1996).

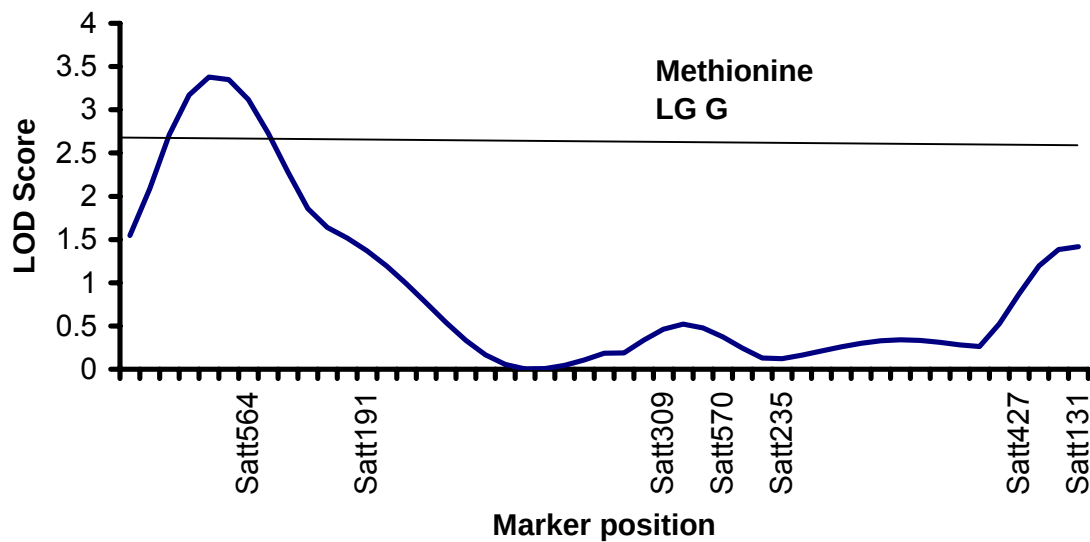


Figure 5.5: LOD score plot on LG G for Methionine QTL. Likely position of Met QTL is upstream of Satt564 on LG G.

It is interesting to note that all QTL associated with Met were also associated with Cys in this population. It should be noted that Met QTL near Satt564 located on LG G was within 6 cM distance from Satt427 (a Cys QTL). An explanation for this observation could be that in the biochemical pathway of Met biosynthesis, Cys is the intermediate product in the process of assimilating sulfur (Matthews, 1999; Saito, 1999). With one amino acid being dependent on the concentration of others, it is possible to detect the same QTL for both Met and Cys.

With this in mind the total sulfur containing amino acid was derived by adding Met and Cys in this population and QTL analysis was performed

on that sum. We detected seven QTL linked to molecular markers Satt002, Satt185, Satt201, Satt235, Satt252, Satt268, Satt427, Satt436 and Satt590 associated with total S-containing amino acids distributed on five molecular linkage groups D2, E, F, G and M (Table 5.3). The significant QTL associated with total S-containing amino acids were the same as those detected either for Cys or Met or both (other than the one near Satt461), as expected.

We know that amino acids including Met and Cys are components of protein. Therefore, QTL associated with Met and Cys are expected to be associated with protein too. In this population, we previously identified some of the QTL associated with nitrogen accumulation in seed at various reproductive growth stages (Panthee et al., 2004). It is interesting to note that not all QTL associated with Met and Cys are associated with total seed protein concentration, although they may be located on the same linkage groups. We found that there was no correlation ($r=-0.13$) between Met and protein whereas the correlation was moderately negative between Cys and protein ($r=-0.48$) (Table 5.2). Thus, despite knowledge that amino acids are the components of protein, there are other genes, which govern total protein concentration in soybean.

Several researchers have mapped QTL for protein in soybean. For example, Brummer et al. (1997) found three QTL on LG G for protein

content. We also found QTL on LG G associated with Met (current study) as well as protein (Panthee et al., 2004), which indicates that LG G contains important QTL for both total seed protein and Met concentration in soybean. Recently, Hyten et al. (2004) mapped protein QTL on various molecular linkage groups including F and M, we detected QTL for Cys and Met on these linkage groups as well.

In quantitative traits, genetic gain or accumulation of favorable alleles in the progeny is possible from both of the parents. In this study QTL analysis helped to reveal the contribution of a particular parent at a particular locus for Cys. The additive genetic effect was assigned with respect to P1, thus sign was positive, the genetic gain was through P1, and if the sign was negative, the genetic gain was through P2 (Table 5.3). Favorable alleles are distributed in the population, which can be brought together through breeding efforts to make significant genetic gains. The QTL for Met and Cys that we report provide a first step to utilize a marker-assisted selection (MAS) programs achieve genetic gains.

Although there are several papers on QTL mapping for other traits, there is very limited information on QTL mapping for amino acids in any species. Loudelt et al. (2003) mapped the QTL for amino acids in *Arabidopsis* and Wang et al. (2001b) mapped the elongation factor involved in lysine biosynthesis in *Zea mays*. Wang et al. (2001a) also mapped a QTL

involved in determining the free amino acid content in *Zea mays*. No papers on QTL mapping for amino acid composition in soybean are available. QTL detected for Met and Cys in the present study have potentially important applications in improving protein quality in soybean by MAS. The QTL detected in the present population need to be verified in other populations to confirm for example, whether conserved alleles for Met and Cys exist among important breeding population parents, or whether specific lines would better facilitate MAS. The recently registered improved S-containing amino acid germplasm TN04-5363 is available to breeders and other researchers aiming to improve the nutritional quality in soybean.

References

- Altenbach, S.B., K.W. Pearson, F.W. Leung, and S.S. Sun. 1987. Cloning and sequence analysis of a cDNA encoding a Brazil nut protein exceptionally rich in methionine. *Plant Mol Biol* 8:239-250.
- Altenbach, S.B., K.W. Pearson, G. Meeker, L.C. Staraci, and S.S. Sun. 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 Mol Biol* 13:513-522.
- Bezant, J.H., D.A. Laurie, N. Pratchett, J. Chojecki, and M.J. Kearsey. 1997. Mapping of QTL controlling NIR predicted hot water extract

- and grain nitrogen content in a spring barley cross using marker-regression. *Plant Breeding* 116:141-145.
- 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., T.E. Carter, and R.F. Wilson. 1999. Registration of Prolina soybean. *Crop Sci* 39:294-295.
- Churchill, G.A., and R.W. Doerge. 1994. Empirical threshold values for quantitative trait mapping. *Genetics* 138:963-971.
- Clarke, E.J., and J. Wiseman. 2000. Developments in plant breeding for improved nutritional quality of soybeans I. Protein and amino acids content. *J Agric Sci* 134:111-124.
- Concibido, V.C., B. La Vallee, P. McLaird, N. Pineda, J. Meyer, L. Hummel, J. Yang, K. Wu, and X. Delannay. 2003. Introgression of a quantitative trait locus for yield from Glycine soja into commercial soybean cultivars. *Theor Appl Genet* 106:575-582.
- Cregan, P.B., T. Jarvik, A.L. Bush, R.C. Shoemaker, K.G. Lark, A.L. Kahler, N. Kaya, T.T. VanToai, D.G. Lohnes, and J. Chung. 1999. An integrated genetic linkage map of the soybean genome. *Crop Sci* 39:1464-1490.
- Falconer, D.S., and T.F.C. Mackay. 1996. *Introduction to Quantitative Genetics*. Fourth ed. Longman Group Ltd.

- Fontaine, J. 2003. Amino acid analysis of feeds, p. 15-40, *In* J. P. F. D'Mello, ed. Amino Acids in Animal Nutrition, Second ed. CABI Publishing, Cambridge, MA.
- Fontaine, J., J. Horr, and B. Schirmer. 2001. Near-infrared reflectance spectroscopy enables the fast and accurate prediction of the essential amino acid contents in soy, rapeseed meal, sunflower meal, peas, fishmeal, meat meal products, and poultry meal. *J Agric Food Chem* 49:57-66.
- Friedman, M., and D.L. Brandon. 2001. Nutritional and health benefits of soy proteins. *J Agric Food Chem* 49:1069-1086.
- George, A.A., and B.O. de Lumen. 1991. A novel methionine-rich protein in soybean seed: identification, amino acid composition, and N-terminal sequence. *J Agric Food Chem* 39:224-227.
- Hyten, D.L., V.R. Pantalone, C.E. Sams, A.M. Saxton, D. Landau-Ellis, T.R. Stefaniak, and M.E. Schmidt. 2004. Seed quality QTL in a prominent soybean population. *Theor Appl Genet* 109:552-561.
- Imsande, J. 2001. Selection of soybean mutants with increased concentrations of seed methionine and cysteine. *Crop Sci* 41:510-515.
- Imsande, J., and J.M. Schmidt. 1998. Effect of N source during soybean pod filling on nitrogen and sulfur assimilation and remobilization. *Plant Soil* 202:41-47.

- Kearsey, M.J., and H.S. Pooni. 1996. The Genetical Analysis of Quantitative Traits. First ed. Chapman and Hall, New York.
- Kortt, -A.A., and -J.B. Caldwell. 1990. Low molecular weight albumins from sunflower seed: identification of a methionine-rich. *Phytochemistry* 29:2805-2810.
- Kwanyuen, P., R.F. Wilson, and J.W. Burton. 1998. Soybean protein quality, p. 284-289, *In* S. S. Koseoglu, et al., eds. Oilseeds and edible oils processing, Vol. vol I. AOCS Press, Champaign, IL.
- Kwanyuen, P., V.R. Pantalone, J.W. Burton, and R.F. Wilson. 1997. A new approach to genetic alteration of soybean protein composition and quality. *J Am Oil Chem Society* 74:983-987.
- Lincoln, S.E., M.J. Daly and E.S. Lander. 1993. Constructing genetic linkage maps with MAPMAKER Version 3.: A tutorial and reference manual. Release 3rd edition. Whitehead Institute for Biomedical Research, Cambridge, MA.
- Liu, S.M., and D.G. Masters. 2003. Amino acid utilization for wool production, p. 309-328, *In* J. P. F. D'Mello, ed. Amino Acids in Animal Nutrition, Second ed. CABI Publishing, Cambridge, MA.
- Loudelt, O., S. Chaillou, P. Merigout, J. Talbotec, and F. Daniel-Vedele. 2003. Quantitative trait loci analysis of nitrogen use efficiency in *Arabidopsis*. *Plant Physiol* 131:345-358.
- Luck, P., T. Lanier, C. Daubert, R.F. Wilson, and P. Kwanyuen. 2001. Functionality and viscoelastic behavior of Prolina soybean isolate, p.

- 197-202, *In* R. F. Wilson, ed. Oilseed processing and utilization. AOCS Press, Champaign, IL.
- Matthews, B.F. 1999. Lysine, Threonine, and Methionine biosynthesis, p. 205-225, *In* B. K. Singh, ed. Plant amino acids: Biochemistry and biotechnology. Marcel Dekker Inc, New York.
- Muntz, K., V. Christov, G. Saalbach, I. Saalbach, D. Waddel, T. Pickardt, O. Schieder, and T. Wustenhagen. 1998. Genetic engineering for high methionine grain legumes. *Nahrung* 42:125-127.
- Njiti, V.N., K. Meksem, M.J. Iqbal, J.E. Johnson, M.A. Kassem, K.F. Zobrist, V.Y. Kilo, and D.A. Lightfoot. 2002. Common loci underlie field resistance to soybean sudden death syndrome in Forrest, Pyramid, Essex, and Douglas. *Theor Appl Genet* 104:294-300.
- Nordlee, J.A., S.L. Taylor, J.A. Twonsend, L.A. Thomas, and R.K. Bush. 1996. Identification of a Brazil-nut allergen in transgenic soybeans. *The New England Journal of Medicine* 334:688-692.
- Nyquist, W.E. 1991. Estimation of heritability and prediction of selection response in plant populations. *Critical Reviews in Plant Sciences* 10:235-322.
- Olsen, M.S., T.L. Krone, and R.L. Phillips. 2003. BSSS53 as a donor source for increased whole-kernel methionine in maize: Selection and evaluation of high-methionine inbreds and hybrids. *Crop Science* 43:1634-1642.

- Pantalone, V.R., F.L. Allen, and D. Landau-Ellis. 2003. Registration of 'TN93-99' Soybean germplasm. *Crop Sci* 43:1137.
- Panthee, D.R., V.R. Pantalone, C.E. Sams, A.M. Saxton, D.R. West, and W.E. Rayford. 2004. Genomic regions governing soybean seed nitrogen accumulation. *J Am Oil Chem Society* 81:77-81.
- Pazdernik, D.L., A.S. Killam, and J.H. Orf. 1997. Analysis of amino and fatty acid composition in soybean seed, using near infrared reflectance spectroscopy. *Agron J* 89:679-685.
- Ravanel, S., B. Gakiere, D. Job, and R. Douce. 1998. The specific features of methionine biosynthesis and metabolism in plants. *Proceedings of National Academy of Sciences, USA* 95:7805-7812.
- Rector, B.G., J.N. All, W.A. Parrott, and H.R. Boerma. 1998. Identification of molecular markers linked to quantitative trait loci for soybean resistance to corn earworm. *Theor Appl Genet* 96:786-790.
- Saito, K. 1999. Biosynthesis of Cysteine, p. 267-291, *In* B. K. Singh, ed. *Plant amino acids: Biochemistry and biotechnology*. Marcel Dekker Inc, New York.
- SAS Institute Inc. 2002. SAS guide for personal computers. 9th ed. SAS Institute, Cary, NC.
- Sexton, P.J., N.C. Paek, and R.M. Shibles. 1998. Effects of nitrogen source and timing of sulfur deficiency on seed yield and expression of 11S and 7S seed storage proteins of soybean. *Field Crops Research* 59:1-8.

- Soytech Inc. 2004. Oilseed statistics. Soytech Inc., Bar Harbor, ME.
- Townsend, J.A., and L.A. Thomas. 1994. Factors which influence the *Agrobacterium*-mediated transformation of soybean. *Journal of Cell Biochemistry Suppl.* 18A:X1-014.
- Wang, S, C, .J. Basten and Z.B. Zeng. 2003. Windows QTL Cartographer. Release V2.0. North Carolina State University, Raleigh, NC.
- Wang, X., D.K. Stumpf, and B.A. Larkins. 2001a. Aspartate kinase 2. A candidate gene of a quantitative trait locus influencing free amino acid content in maize endosperm. *Plant Physiol* 125:1778-1787.
- Wang, X., Y. Won, C.S. Kim, and B.A. Larkins. 2001b. Quantitative trait locus mapping of loci influencing elongation factor 1a content in maize endosperm. *Plant Physiol* 125:1271-1282.
- Whitehead, T.F., V.T. Sapro, and D.R. Rao. 1989. Sulfur-containing amino acids and protein content in selected soybean progenies. *Soybean Genetics Newsletter* 16:74-80.
- Wilson, R.F. 1987. Seed metabolism, p. 643-686, *In* J. R. Wilcox, ed. Soybeans: Improvement, production and uses, Second ed. ASA, CSSA, SSSA, Madison, WI.
- Wright, A., and B. Orman. 1995. Rapid screening procedure for methionine levels in maize and soybean. *Crop Sci* 35:585-586.
- Zarkadas, C.G., H.D. Voldeng, Z.R. Yu, and V.K. Choi. 1999. Assessment of the protein quality of nine northern adapted yellow and brown

seed coated soybean cultivars by amino acid analysis. J Agric Food Chem 47:5009-5018.

Zarkadas, C.G., H.D. Voldeng, Z.R. Yu, K. Shang, and P.L. Pattison. 1997. Comparison of the protein quality of five new northern adapted natto soybean cultivars by amino acid analysis. J Agric Food Chem 45:2013-2019.

PART 6

Genomic Regions Associated with Amino Acid Composition in Soybean

This chapter is a lightly revised version of paper by the same name that has been submitted to the ***Molecular Breeding*** in 2004 by D.R. Panthee, V.R. Pantalone, A.M. Saxton, D.R. West and C.E. Sams

My use of “our” in this chapter refers to my co-authors and myself. My primary contribution to this paper include (1) organization, development and management of all aspects of the project, (2) collecting all molecular data using polymorphic molecular markers, (3) collecting all the field data at Knoxville and Holston environments, (4) analyzing for amino acids composition, (5) analyzing all data, and (6) all of the writing except for some minor revisions.

Abstract

Soybean [*Glycine max* (L.) Merr.] is the single largest source of protein in animal feed. However, few studies have been conducted to evaluate genomic regions controlling amino acid composition in soybean. It is important to study the genetics of amino acid composition to achieve improvements through breeding. The objective of this study was to determine genotypic differences in soybean recombinant inbred lines (RIL) and to identify genomic regions controlling amino acid composition in soybean seed. To achieve this objective, 101 F₆-derived RIL from a population developed from a cross of N87-984-16 x TN93-99 were used. Ground soybean seed samples were analyzed for 18 amino acids using a

near infrared spectroscopy (NIRS) instrument. The data were analyzed using SAS software. A significant ($p < 0.05$) difference among the RIL was found for amino acid composition. Heritability estimates on entry mean basis ranged from 0.13 for His to 0.67 for Tyr. A total of 94 polymorphic simple sequence repeat (SSR) molecular genetic markers were screened in DNA from progenies. Single factor ANOVA was used to identify candidate QTL, which were then confirmed by QTL Cartographer. At least one QTL for each amino acid was detected in this population. Some markers (e.g. Satt274, Satt143, Satt102, Satt168 etc.) were associated with several different amino acids suggesting that these genomic regions govern general pathway components of amino acid biosynthesis. Molecular linkage group A1 and L were found to have six QTL each and D1b and G each contained five amino acid QTL. Phenotypic variation explained by an individual QTL ranged from 9.4% to 45.3%. QTL detected for amino acids in soybean in this experiment are expected to be useful for future breeding programs targeting development of improved soybean amino acid composition for human and animal nutrition.

Key words: Amino acid, genomic regions, *Glycine max*, lysine, QTL mapping, soybean, threonine, tryptophan

Introduction

Soybean [*Glycine max* (L.) Merr.] is a major source of protein for humans and livestock throughout the world. It is estimated that more than 70% of the total protein meal comes from soybean (Soytech Inc., 2004). A major function of proteins in nutrition is to supply adequate amounts of required amino acids (Friedman and Brandon, 2001). Amino acids are the principal building blocks of proteins and enzymes.

Depending up on the requirement and availability of amino acids in the animal metabolic processes, all 20 amino acids have been divided into two groups, *essential* and *non-essential*. Non-essential amino acids are readily available or can be synthesized by animals, hence they need not be supplied from external sources, whereas essential amino acids cannot be synthesized in animals but play a crucial role in metabolic processes. The essential amino acids are Lysine (Lys), Histidine (His), Leucine (Leu), Isoleucine (Ile), Valine (Val), Methionine (Met), Threonine (Thr), Tryptophan (Trp) and Phenylalanine (Phe) (D'Mello, 2003). Some nutritionists do not include His as an essential amino acid. Differences in classifying the amino acids as essential or non-essential are based on the type of animal and its nutritional needs. For example, humans can produce 10 of the 20 amino acids whereas swine can produce only nine, others must be supplied in the feed. Failure to obtain an adequate quantity

of even a single essential amino acid leads to degradation of the body's proteins to obtain the deficient amino acid. Unlike fat and starch, the human body does not store excess amino acids for later use, therefore the amino acids must be obtained from food every day.

Although much attention has been given to essential amino acids, non-essential amino acids are also important in developing an ideal protein. Non-essential amino acids are a good source of nitrogen, which is required for the synthesis of essential amino acids and other biochemical compounds. Therefore, it is important to determine the ratio of essential to non-essential amino acids in a feedstuff to determine the idealness of the protein (Heger, 2003). Some researchers have expressed the ratio in terms of nitrogen of essential to non-essential amino acids whereas others have expressed the ratio in terms of nitrogen of essential to total amino acids. Since non-protein nitrogen is close to zero (Heger, 2003), the latter can be expressed as the ratio between nitrogen of essential amino acids to total amino acids (E:T). In this paper, we report the ratio of essential to non-essential, and essential to total amino acids.

Among the essential amino acids, Lys, Met, Trp and Thr are the most widely used in animal feed industry. Traditionally, synthetic forms have been used to fulfill the requirements of the essential amino acids. Fontaine

(2003) estimated that more than 400,000 mt of Lys and Met each are used annually in the world to fulfill the requirement of animal feed annually. Realizing the importance of amino acid composition in soybean, a few of the northern adapted soybean cultivars were analyzed by a High Pressure Liquid Chromatography (HPLC) method (Zarkadas et al., 1995; Zarkadas et al., 1993; Zarkadas et al., 1997a; Zarkadas et al., 1999; Zarkadas et al., 1997b). Later, near infrared reflectance spectroscopy (NIRS) based method was used (Pazdernik et al., 1997) to analyze the amino acid composition of soybean. The NIRS based method is fast, easy and accurate in analysis (Fontaine, 2003).

Despite the central role of amino acids in animal nutrition, little information is available about the genetic control of amino acids in soybean. It is important to understand the genetics of a trait in detail to optimize its improvement through breeding. Such efforts in corn (*Zea mays* L) have been made towards the understanding of genetics of lysine (Wang and Larkins, 2001; Wang et al., 2001a; Wang et al., 2001b). However, no such studies have been conducted in soybean. Objectives of the present study were to determine the E:NE and E:T ratio in soybean, and to identify the genomic regions controlling amino acid composition in soybean.

Materials and Methods

Plant materials

A total of 101 F_6 -derived recombinant inbred lines (RIL) were developed from a cross of N87-984-16 x TN93-99. The N87-984-16 line is one of two sister lines, whose blend constitutes the high protein commercial cultivar 'Prolina' (Burton et al., 1999). The line TN93-99 is a high yielding and well adapted University of Tennessee breeding line and is currently registered as a germplasm GP-280 (Pantalone et al., 2003). Crossing was made during summer of 1998. F_1 seeds were harvested in October 1998 and F_2 single plants were grown in Costa Rica during winter 1998. Generations were advanced via single seed descent until the F_5 in Costa Rica and $F_{5.6}$ seeds were obtained in May 2000. Approximately 300 $F_{5.6}$ single plants were grown at Knoxville as a recombinant inbred line (RIL) population and 101 random plants of similar Tennessee adapted maturity were selected as a source population for this study. All 101 lines and parents were planted in 6 m length two-row plots with three replications in a randomized complete block design (RCBD) at Knoxville Plant Science Farm (KPSF) in 2001. All the RIL, the two parents, as well as three checks (Hutcheson, 5002T and 5601T) were planted in a RCBD with three replications at three locations at the Knoxville and the Holston units of the Knoxville Experiment Station of the University of Tennessee, and at Ames Plantation Experiment Station, in Grand Junction, TN in May 2002 and

2003. Each line was planted in a four row plot of 6 m length with a spacing of 75 cm between rows.

Sample preparation and amino acid analyses

Approximately 20 g of soybean seeds were ground in a water-cooled Knifetec 1095 Sample Mill (FOSS Tecator, S-26321, Hogana, Sweden) for 20 seconds. This setting produced soybean flour with a uniform particle size. The NIR instrument (NIR 6500, FOSS North America) was warmed up for 2 h after turning on the lamp, then auto diagnostics were run for instrument response, wavelength accuracy and NIR repeatability. A dehumidifier was used throughout the analysis setting the humidity to 40%, and room temperature was approximately 20°C. The ground soybean samples were scanned using Winisi II 1.5 version of software. The instrument was left on for the whole period of analyses and diagnostics was performed every day until the scanning was finished. Sample scanning produced the predicted concentration of amino acids.

DNA extraction and polymerase chain reaction

DNA was extracted from the RIL and parental lines utilizing Qiagen Plant Easy DNA Extraction Kit (Qiagen Hilden, Germany). Polymerase chain reaction (PCR) consisted of 7.4 µL of ddH₂O, 1 µL of 10X PCR Buffer, 1 µL of 2 mM dNTPs mixture (Pharmacia, Piscataway, NJ), 0.5 µL of 20 µM

forward and reverse primer, 0.1 µL of 5 units/µL KlenTaq (Ab Peptides Inc., St. Louis, MO) and 2 µL of 20 ng/µL template DNA. The PCR was performed in a 96-well MBS Hybaid thermocycler (Hybaid, Franklin, MA). PCR conditions were a) 94°C for five minutes followed by 35 cycles at 94°C for denaturation for 25 s, b) 47°C for annealing for 30 s, c) 72°C for 25 s for extension, and one last cycle at 72°C for final extension for five minutes. Parents were screened for a total of 585 (ATT)_n type of simple sequence repeat (SSR) genetic markers developed by Cregan *et al.* (1999), of which 138 were polymorphic. The sequence information for the markers is publicly available from the Website (www.soybase.org), which was verified on December 16, 2004. Out of 138, only 94 polymorphic markers could be scored for QTL analysis. The remaining 44 either did not amplify in the progeny or could not be scored because of indistinguishable band pattern.

DNA gel electrophoresis

A 6% non-denaturing polyacrylamide gel electrophoresis (PAGE) used to separate the PCR product consisted of 6% bis-acrylamide, 0.5% TBE buffer, 0.07% APS and 0.035% TEMED, which was measured as 28.5 mL of 40% bis-acrylamide, 160.2 mL of 0.593X TBE buffer, 1.33 mL of 10% APS and 66.5 µL of TEMED. Two µL of loading buffer (6X) were added to PCR products and a 10 µL sample was loaded on the gel. The running

buffer was 0.5X TBE. The gel was run at constant 300 V for 3 h. A fan was used to circulate air to keep the glass plates cool during gel running. Ethidium bromide (50 µl, 10 mg/mL) was added to the running buffer in order to visualize the bands under exposure to UV light. Bands were scored using 1 to represent P₁ (N87-984-16), 2 to represent heterozygote, and 3 to represent P₂ (TN93-99) alleles for each primer locus.

Data analysis

Phenotypic data on amino acids were analyzed using the MIXED procedure of SAS software to determine the genotypic differences among the RIL. Location and replication were considered as two blocking factors in the model. Heritability of the trait in the population was estimated on an entry mean basis (Nyquist, 1991) as follows:

$$h^2 = \frac{\sigma_g^2}{\sigma_g^2 + \left(\frac{\sigma_{ge}^2}{e} \right) + \left(\frac{\sigma^2}{re} \right)}$$

where, h^2 represents the heritability, σ_g^2 is genotypic variance, σ_{ge}^2 is genotype x environment variance, σ^2 is error variance, r is number of replications and e is the number of environments.

PROC MIXED COVTEST was used to estimate variance components for calculating heritability estimates. Phenotypic correlation was determined

using the CORR procedure. Associations between molecular data and least square means of NIRS predicted amino acids were detected by analyzing the data using the GLM procedure (SAS Institute Inc, 2002). The heterozygotes were excluded from the analysis because the RIL population was F₆-derived, and our interest was to detect heritable alleles from pure lines. QTL position and distance were estimated using Mapmaker/Exp 3.0 (Lincoln et al., 1993) and the information from Mapmaker was then used in QTL Cartographer (Wang et al., 2003) to confirm the QTL by composite interval mapping. We used standard model *Zmapqtl* 6 with 10 cM window size and 2 cM walking speed. We performed 1000 permutations on all amino acids in each environment and in the combined data from six environments to establish the empirical LOD thresholds at the 5% probability level (Churchill and Doerge, 1994).

Results and Discussion

Phenotypic traits

There was a significant difference ($p < 0.01$) among the RIL for various essential and non-essential amino acids in the soybean population. Despite the significant difference among the RIL, the range was narrow for some of the amino acids such as Ser, Thr, Gly, Ala, Ile, Tyr and Trp (Table 6.1). However, there was a reasonable level of variation for other amino acids. Low variation suggests that it may be difficult to make significant

Table 6.1: Descriptive statistics (g kg⁻¹ seed dry weight) of various amino acid and total protein concentration in soybean seed in an F6-derived population of N87-984-16 x TN93-99.

Trait	Minimum	Maximum	Mean	Std. Dev.	P1	P2	LSD _{0.05}	h ² †
Asp	43.0	47.0	45.0	0.8	46.9	43.5	0.04	0.57
Glu	56.2	64.4	60.3	1.7	64.4	57.2	0.11	0.52
Ser	26.8	28.9	27.9	0.5	28.0	27.4	0.04	0.20
Thr	19.6	20.6	20.2	0.2	20.0	20.1	0.02	0.23
Gly	29.7	32.0	30.9	0.5	30.6	30.7	0.05	0.34
Ala	22.5	23.6	23.0	0.3	22.9	22.9	0.02	0.38
Pro	23.4	25.2	24.4	0.3	24.7	24.1	0.02	0.55
Val	26.6	28.4	27.6	0.4	27.5	27.5	0.03	0.63
Arg	34.9	38.4	37.1	0.7	37.1	36.9	0.03	0.54
Ile	19.3	20.7	20.0	0.3	20.4	19.6	0.01	0.57
Leu	26.5	30.0	28.3	0.7	29.3	27.2	0.04	0.60
Phe	20.7	23.2	22.0	0.4	23.0	21.2	0.03	0.38
Lys	22.4	24.1	23.3	0.4	24.1	22.9	0.03	0.38
His	20.1	22.9	21.6	0.6	22.4	21.0	0.05	0.13
Tyr	19.3	20.9	20.4	0.3	20.0	20.3	0.02	0.67
Trp	4.4	5.1	4.8	0.1	5.1	4.6	0.01	0.50
Cys	5.1	7.3	6.2	0.4	7.0	6.0	0.03	0.14
Met	4.4	8.8	6.4	0.9	6.8	6.8	0.07	0.57
Protein	399.7	434.5	416.1	7.0	431.0	401.0	0.33	0.66

P1 = N87-984-16; P2 = TN93-99; LSD_{0.05} = Least significant difference at

the 0.05 probability level; †heritability on an entry mean basis over six environments

genetic improvements by selecting a line for a particular amino acid in this population. However, gradual improvements remain possible by selecting genotypes from upper tail of the distribution for other amino acids. Repeated selection followed by inter-mating for a particular amino acid would eventually be helpful in increasing the frequencies of desirable alleles. The majority of the essential amino acids such as Val, Leu, Phe, Lys, His and Met had reasonably good variation, indicating that genetic gains by means of selection is likely.

The heritability estimates for the amino acids was low to moderately high, a reflection of genetic variation (Table 6.1). In this population, Asp, Glu, Pro, Val, Arg, Ile, Leu, Tyr, Trp and Met had moderately high heritability estimates whereas the remaining amino acids had low to medium. Interestingly, Thr, Met, Leu, Ile, Phe, Trp and Val are essential amino acids (D'Mello, 2003), which had medium to high levels of heritability again suggesting that genetic improvements could be attainable. It has been established on the basis of past research that Lys, Thr, Met and Trp, in order of importance, are crucial in swine diets (Boisen, 2003). The most limiting five essential amino acids for young poultry are Lys, Trp, Arg, Thr and Val (Baker, 2003).

Considering the variation available in the present soybean population, heritability estimates and amino acid requirements for swine and poultry, there is a possibility of gradual genetic improvement for Thr, Val, Met, Leu and Phe. It is expected that selection response in a population for these amino acids should be attainable.

The low phenotypic variation found in the present population for some of the amino acids is probably because of limited contrasts between the two parents for amino acid concentration (Table 6.1). In fact there are very few papers on genetic improvement of amino acid composition in soybean. A few northern soybean cultivars have been assessed for amino acid composition (Zarkadas et al., 1993; Zarkadas et al., 1997a; Zarkadas et al., 1999; Zarkadas et al., 1997b; Zarkadas et al., 1994) but there are virtually no reports from the southern gene pool. The findings from the present experiment are expected to provide some basis for future research in genetic improvement of soybean for amino acid composition.

The least number of amino acids classified as essential for all animal species is eight (Lys, Leu, Ile, Val, Met, Thr, Trp and Phe). For the purpose of discussion, we have denoted these amino acids as 'basic essential amino acids'. Depending upon the requirements of animal

species, more amino acids have been added into this list as an essential. While computing a balance diet for a particular animal species, one of the approaches adopted is to determine the ratio between essential to non-essential or essential to total amino acids (Heger, 2003). We determined this ratio considering various amino acids as essential so that we can get information about the suitability of soybean for a particular animal species. Furthermore, it would also be helpful in determining the addition of deficient amino acids to obtain the desirable ratio in animal feeds. We have reported both the essential to non-essential and the essential to total amino acid ratio (Table 6.2). However, for the convenience of discussion in relation to the available literature, we will be using essential to total amino acid ratio in this paper.

Unlike individual amino acid concentration, there was no variation for essential to total amino acids ratio, which averaged 0.34 for basic essential amino acids in our population. When His was included as an essential amino acids, the average ratio increased to 0.39, and when Arg, Cys and Tyr were also included, the average ratio further increased to 0.53 (Table 6.2). It is obvious that inclusion of various amino acids makes the ratio improve, which has been described in length by Heger (2003).

Table 6.2: Ratio between essential to non-essential (E:NE), and essential to total (E:T) amino acids in an F6-derived population of soybean from N87-984-16 x TN93-99.

Descriptors	Total Amino	Essential	Non- essential	E:T	E:NE
Basic 8 essential amino acids†					
Minimum	43.7	14.9	28.8	0.34	0.51
Maximum	46.2	15.6	30.6	0.34	0.52
Mean	45.0	15.3	29.7	0.34	0.51
Std. Dev.	0.54	0.15	0.40	0.00	0.00
Basic 8 essential amino acids +His					
Minimum	43.7	16.9	26.7	0.38	0.62
Maximum	46.2	17.9	28.3	0.39	0.65
Mean	44.9	17.4	27.5	0.39	0.63
Std. Dev.	0.54	0.19	0.38	0.00	0.01
Basic 8 essential amino acids +His+Arg+Cys+Tyr					
Minimum	43.7	23.1	20.4	0.52	1.10
Maximum	46.2	24.4	21.8	0.53	1.15
Mean	44.9	23.8	21.2	0.53	1.12
Std. Dev.	0.54	0.24	0.32	0.00	0.01

E:T = Ratio of essential to total amino acids concentration.

E:NE = Ratio of essential to non-essential amino acids concentration

†The basic eight essential amino acids are Lys, Leu, Ile, Val, Met, Thr, Trp and Phe

The observed ratio in the present RIL population of soybean was close to that found optimum for swine and poultry (0.60) by including His, Arg, Cys and Tyr as an essential amino acids (Heger, 2003). Therefore, the average soybean meal and certainly that from selected superior RIL from this population seems ideal from an amino acids composition point of view for swine and poultry.

Phenotypic correlations among amino acids in the present soybean population revealed that there was no clear pattern of correlation between essential and non-essential amino acids, rather correlations were based on individual amino acid (Table 6.3). For example, Lys, an important essential amino acid, was negatively correlated with Gly whereas Lys was positively correlated with Glu, both of which are non-essential amino acids. Similarly, Lys was negatively correlated with Val whereas Lys was positively correlated with Leu, which are essential amino acids (Table 6.3). This suggests that while screening genotypes in an improvement program, it is important to assess the genotype for all individual amino acids since there is no clear pattern of relationship between essential and non-essential amino acids. Furthermore, at least four amino acids had low or no positive correlations with protein, indicating that high protein genotypes may not necessarily have high amino acids.

Table 6.3: Simple phenotypic correlation coefficients between amino acids in soybean seeds in an F6-derived population of N87-984-16 x TN93-99.

	Asp	Thr	Ser	Glu	Pro	Gly	Ala	Cys	Val	Met	Ile	Leu	Tyr	Phe	His	Lys	Arg	Trp
Thr	-0.26																	
Ser	-0.19	0.94																
Glu	0.80	-0.47	-0.47															
Pro	0.51	0.62	0.63	0.17														
Gly	-0.09	0.57	0.66	-0.50	0.48													
Ala	0.47	0.32	0.35	0.05	0.78	0.68												
Cys	-0.29	0.58	0.60	-0.24	0.13	-0.10	-0.41											
Val	0.30	0.53	0.55	-0.07	0.79	0.75	0.90	-0.22										
Met	0.16	0.28	0.26	0.02	0.41	-0.11	0.12	0.41	-0.05									
Ile	0.86	-0.33	-0.30	0.71	0.42	0.07	0.62	-0.64	0.47	-0.13								
Leu	0.50	-0.54	-0.54	0.72	-0.14	-0.31	-0.01	-0.50	-0.01	-0.57	0.64							
Tyr	-0.23	0.68	0.66	-0.41	0.40	0.69	0.40	0.07	0.67	-0.30	-0.07	-0.12						
Phe	0.04	0.62	0.64	0.06	0.42	-0.01	-0.17	0.85	0.00	0.41	-0.32	-0.28	0.20					
His	-0.28	0.80	0.82	-0.36	0.38	0.24	-0.08	0.90	0.10	0.51	-0.55	-0.64	0.31	0.83				
Lys	0.29	-0.46	-0.36	0.43	-0.29	-0.56	-0.52	0.26	-0.55	0.03	0.00	0.27	-0.52	0.21	-0.02			
Arg	-0.18	0.44	0.48	-0.19	0.20	0.39	0.03	0.32	0.31	-0.23	-0.16	-0.03	0.49	0.39	0.44	0.11		
Trp	0.21	0.77	0.78	0.05	0.71	0.23	0.21	0.68	0.35	0.44	-0.05	-0.24	0.37	0.83	0.77	-0.07	0.32	
Prt	0.39	0.63	0.62	0.26	0.72	0.19	0.29	0.53	0.37	0.40	0.16	-0.03	0.30	0.74	0.61	-0.05	0.20	0.92

Prt = Total protein

In the past, attempts have been made to assess the protein quality in response to soil nutrition (Lhuillier-Soundele et al., 1999; Paek et al., 2000; Sexton et al., 1998). These researchers have analyzed protein fractions to show the relationships between externally supplied nutrients and protein or protein fractions but not the amino acids. Furthermore, it has been suggested that there may be a role for nutrients in the expression of genes related to protein quality in soybean (Fujiwara et al., 1992; Kim et al., 1999; Nakasathien et al., 2000).

QTL analysis

We initially identified numerous QTL for amino acids on the basis of single factor ANOVA and then confirmed many of them by composite interval mapping (CIM) using QTL Cartographer (Wang et al., 2003). Only QTL that were either: a) consistent across the environments and significant at $LOD \geq 2.0$ via QTL Cartographer or b) environmentally sensitive yet significant at empirically determined LOD threshold levels are reported in this paper. Because this paper is the first report of soybean amino acid concentration, we feel that this compromise represents the most balanced approach to guide future research.

Only one QTL near Satt380 for His was detected in this population, the increased allele was contributed through P2. That single QTL explained 16.8% of total phenotypic variation indicating that it was a major QTL (Falconer and Mackay, 1996) associated with His. There were five QTL linked to molecular markers Satt168, Satt177, Satt203, Satt252 and Satt313 associated with Ile located on linkage groups A2, B2, D1a, F and L in this population (Table 6.4). Phenotypic variation explained by the Ile QTL ranged from 10.6 to 15.7%. In three of the five loci, favorable alleles for increasing Ile were contributed through P2 although P1 was the high parent for this trait (Table 6.1 and 6.4).

Table 6.4: Quantitative trait loci associated with essential amino acids in soybean seed in an F6-derived population of N87-984-16 x TN93-99.

Trait	LG	Marker	Position (cM)	LOD Score	Additive Effect#	R ² (%)
His	J	Satt380	43.11	2.80‡	-2.88	16.8
Ile	A2	Satt177	36.77	2.52‡	-1.54	12.7
	B2	Satt168	55.20	2.72‡	0.01	10.6
	D1a	Satt203	59.00	2.98 [†]	1.64	14.0
	F	Satt252	16.08	3.25 [†]	-1.80	15.7
	L	Satt313	34.54	2.64‡	-0.01	13.6
Leu	D1a	Satt203	59.00	3.39 [†]	0.02	17.7
	D1b	Satt274	116.35	3.15 [†]	0.02	17.3
	D2	Satt002	47.73	2.89‡	0.03	16.4
	G	Satt570	12.74	2.83‡	0.02	13.6
	L	Satt076	61.35	2.95 [†]	0.02	13.4
	M	Satt540	35.85	2.71‡	0.03	14.3
	M	Satt567	33.47	2.64‡	0.03	10.4
	N	Satt255	76.49	3.14 [†]	0.02	12.3
Lys	D1a	Satt184	17.52	3.11 [†]	0.01	13.1
	E	Satt185	44.76	3.93 [†]	-0.01	14.7
	E	Satt268	44.27	3.24 [†]	-0.02	11.8
	G	Satt235	21.89	3.07 [†]	0.02	10.8
	G	Satt427	51.69	3.62 [†]	2.25	17.3
Phe	B2	Satt168	55.20	3.39 [†]	0.01	12.8
	D1b	Satt274	116.35	3.56 [†]	-0.01	13.9
	J	Satt249	12.33	3.17 [†]	-3.22	29.1
	J	Satt380	43.11	3.08 [†]	-2.15	12.7
	K	Satt102	30.28	2.74‡	0.01	9.7
	L	Satt495	0.00	3.36 [†]	-0.01	15.7
Thr	A1	Satt593	25.56	2.99 [†]	-0.01	17.0
	D1b	Satt274	116.35	2.82‡	-1.11	9.4
	K	Satt518	46.63	2.83‡	-1.22	21.3
	L	Satt143	30.19	3.07 [†]	-0.01	11.8
	L	Satt278	31.22	3.37 [†]	-0.01	13.0
	L	Satt495	0.00	4.00 [†]	-0.01	45.3
Trp	A1	Satt236	93.23	2.54‡	-0.44	10.4
	C1	Satt139	74.46	2.67‡	0.01	18.2
	D1a	Satt203	59.00	3.20 [†]	0.56	12.8
	D1b	Satt537	75.67	2.83‡	0.01	16.3
	G	Satt570	12.74	2.91‡	0.51	10.4
	I	Satt292	82.78	3.10 [†]	0.72	23.0
Val	B2	Satt168	55.20	2.58‡	0.03	10.7
	D1b	Satt600	75.41	2.71‡	-0.03	10.5
	F	Satt252	16.08	4.45 [†]	-0.02	20.5
	L	Satt143	30.19	3.37 [†]	-2.88	16.3
	L	Satt278	31.22	2.87‡	-2.27	10.2

LG = Linkage group; [†]Significant at empirical LOD threshold level determined by 1000 permutations; ‡QTL are significant at default LOD score of 2.0, which is $P \leq 0.01$, $\# \frac{1}{2}(P1-P2)$.

Among the eight QTL linked to molecular markers Satt002, Satt076, Satt203, Satt255, Satt274, Satt540, Satt567 and Satt570 associated with Leu, QTL near Satt540 and Satt567 located on LG M are within 2.4 cM distance of each other indicating that both of these molecular markers represent a single genomic region associated with Leu (Table 6.4). QTL near Satt203 located on LG D1a was the only common QTL between Ile and Leu. It was interesting to note that all the alleles for increased Leu were contributed through P1.

There were five QTL linked to molecular markers Satt184, Satt185, Satt235, Satt268 and Satt427 located on three linkage groups associated with Lys (Table 6.4). QTL near Satt185 and Satt268 located on LG E are within 0.50 cM of each other indicating that both markers represent the same genomic region associated with Lys. Increasing alleles in this genomic region were contributed through P2 whereas at remaining loci, they were contributed through P1. There were two different QTL near Satt235 and Satt427 located almost 40 cM apart on LG G associated with Lys. Increasing alleles at both the loci were contributed through P2. Hence molecular linkage group E and G were found to contain important genomic region associated with Lys in soybean seed.

There were six QTL linked to molecular markers Satt102, Satt168, Satt249, Satt274, Satt380 and Satt495 associated with Phe distributed

among five molecular linkage groups (Table 6.4). Two QTL near Satt249 and Satt380, located on LG J were within 31 cM distance, indicating that they may be within the same genomic region.

Favorable alleles at both of these loci were contributed through P2 and magnitude of contribution from this region of LG J was greater than that from other loci. The molecular LG J may have an important genomic region associated with Phe in soybean seed.

Six QTL linked to molecular markers Satt143, Satt274, Satt278, Satt495, Satt518 and Satt593 associated with Thr. Three of them near Satt143, Satt278 and Satt495 are within 31 cM distance on LG L, and two of them are only 1.03 cM apart. All three of these molecular markers may be indicative of a single QTL, all of those shared a positive contribution through P2. Satt495 positioned at 0.0 cM at the distal end of LG L explained more than 45% of the total phenotypic variation, thus it is an important major QTL for Thr. Genomic region near Satt274, in addition to Leu and Phe QTL, was also associated with Thr indicating that there may be an important QTL involved in the biosynthesis pathway of general amino acids (Table 6.4).

There were six QTL linked to molecular markers Satt139, Satt203, Satt236, Satt292, Satt537 and Satt570 associated with Trp distributed

among six molecular linkage groups (Table 6.4). Among those, genomic region near Satt203 was common with Ile, Leu and Trp. QTL near Satt292 was the strongest QTL for Trp explaining 23% of total phenotypic variation.

There were five QTL linked to molecular markers Satt143, Satt168, Satt252, Satt278 and Satt600 associated with Val located on four molecular linkage groups (Table 6.4). Molecular markers Satt143 and Satt278 on LG L are within 1.03 cM distance indicating that this is a single genomic region governing a QTL associated with Val. In this region, the increasing allelic contribution was through P2 and the extent of the contribution was greater than at other loci. LOD score, allelic contribution and coefficient of determination in this genomic region indicated that LG L contains a major QTL for Val in soybean seed.

Among the essential amino acids Ile, Leu and Val, called branched-chain amino acids, have a common biosynthesis pathway but none of them are interdependent (Singh, 1999). We detected one common QTL between Ile and Leu, and three between Ile and Val.

There is enormous interest of animal feed industries and nutritionists in essential amino acids mainly to obtain the desired concentration of a particular amino acid to optimize feed formulations. This is particularly

important when feed industries are using synthetic amino acids (Clarke and Wiseman, 2000) to obtain the desired ratios. The QTL identified in our soybean population may be useful for obtaining amino acids by natural means. To be conservative, we report only QTL with association in two or more environments. We assume that those QTL should be more stable (Brummer et al., 1997). However, it is always necessary to validate the putative QTL in more than one environment and population before using them in marker-assisted selection or for positional cloning programs. We encourage other researchers to pursue those opportunities.

Among the non-essential amino acids, only one QTL near Satt274 was detected for Arg, and it explained 10.8% of total phenotypic variation. Considering the limited phenotypic variation explained by the QTL, we assume that there may be more QTL associated with Arg, however we could not detect in this population. There were four QTL each for Asp (near Satt177, Satt196, Satt203 and Satt292) and Pro (near Satt143, Satt168, Satt437 and Satt570) in this population (Table 6.5). QTL near Satt177 was the strong one associated with Asp having a high LOD score at 3.4 and explaining 26.1% of the total phenotypic variation for that amino acid. QTL near Satt292 was also a major one associated with Asp ($R^2=34.1\%$) (Table 6.5). All four of the Pro QTL were major QTL and they were each located on different linkage group.

Table 6.5: Quantitative trait loci associated with non-essential amino acids in soybean seed in an F6-derived population of N87-984-16 x TN93-99.

Trait	LG	Marker	Position (cM)	LOD Score	Additive Effect#	R ² (%)
Ala	A1	Satt042	27.66	2.75‡	-0.01	11.2
	A2	Satt437	107.05	3.28†	-1.74	13.4
	B2	Satt168	55.20	3.00†	0.02	11.9
	F	Satt252	16.08	4.11†	-0.01	18.2
	L	Satt143	30.19	5.21†	-0.01	23.5
	L	Satt278	31.22	4.27†	-2.48	15.7
	L	Satt418	30.93	2.53‡	-0.01	10.7
	L	Satt495	0.00	3.21†	-1.46	22.1
Arg	D1b	Satt274	116.35	2.68‡	-2.42	10.8
Asp	A2	Satt177	36.77	3.37†	-0.55	26.1
	D1a	Satt203	59.00	2.71‡	0.42	14.7
	I	Satt292	82.78	2.75‡	5.47	34.1
	K	Satt196	104.79	2.97†	0.56	18.8
Glu	A2	Satt177	36.77	3.13†	-0.09	22.2
	D1a	Satt203	59.00	2.55‡	0.81	13.8
	D1b	Satt274	116.35	3.73†	0.06	15.3
	J	Satt249	12.33	2.97†	0.06	12.9
	L	Satt076	61.35	4.62†	0.06	24.5
	M	Satt567	33.47	3.11†	0.06	15.3
	O	Satt420	49.71	3.57†	-0.04	13.0
	O	Satt479	54.20	3.36†	-0.65	12.2
Gly	A1	Satt042	27.66	2.82‡	-0.02	10.3
	A1	Satt593	25.56	2.91‡	-3.03	16.2
	D1b	Satt274	116.35	3.11†	-2.55	11.5
	K	Satt518	46.63	4.15†	-0.03	23.2
	L	Satt143	30.19	2.80‡	-2.26	11.8
	L	Satt495	0.00	3.83†	-0.83	36.5
Pro	A2	Satt437	107.05	3.26†	-2.13	15.5
	B2	Satt168	55.20	3.04†	0.03	12.3
	G	Satt570	12.74	2.66‡	1.51	11.2
	L	Satt143	30.19	3.64†	-2.10	15.6
Ser	A2	Satt437	107.05	2.84‡	-2.86	13.3
	C1	Satt139	74.46	3.04†	3.12	24.3
	K	Satt102	30.28	3.19†	2.20	12.5
	L	Satt143	30.19	5.41†	-0.02	23.2
	L	Satt495	0.00	4.92†	-0.53	26.9
	M	Satt567	33.47	4.09†	-2.53	15.7
Tyr	A2	Satt133	125.38	3.67†	-0.01	22.9
	D1b	Satt274	116.35	3.51†	-1.66	12.5
	E	Satt185	44.76	7.44†	2.27	23.9
	E	Satt268	44.27	3.87†	0.02	13.4
	K	Satt102	30.28	3.49†	0.01	11.9
	L	Satt143	30.19	2.96†	-1.58	12.1
	L	Satt313	34.54	3.23†	-1.72	14.5
	L	Satt418	30.93	3.07†	-1.42	9.8

LG = Linkage group; † = Significant at empirical LOD threshold level determined by 1000 permutations.; ‡ = QTL are significant at default LOD score of 2.0, which is $P \leq 0.01$, # = $\frac{1}{2}(P1-P2)$.

Six QTL each were detected for Gly (near Satt042, Satt143, Satt274, Satt495, Satt518 and Satt593) and six for Ser (near Satt102, Satt139, Satt143, Satt437, Satt495 and Satt567) in this population (Table 6.5). QTL near Satt495 was the strongest QTL associated with Gly and Ser explaining 36.5% and 26.9% of total phenotypic variation for those amino acids respectively. QTL near Satt143 was a common QTL between Gly and Ser (Table 6.5). The amino acids Gly and Ser has a common biosynthesis pathway; Gly is obtained through Ser (Bourguignon et al., 1999). In the present study, we detected a major common genomic region on LG L associated with both of these amino acids suggesting of their pathway regulation.

There were eight QTL each associated with Ala (near Satt042, Satt143, Satt168, Satt252, Satt278, Satt418, Satt437 and Satt495), Glu (near Satt076, Satt177, Satt203, Satt249, Satt274, Satt420, Satt479 and Satt567) and Tyr (near Satt102, Satt133, Satt143, Satt185, Satt268, Satt274, Satt313 and Satt418) in this population (Table 6.5). Three QTL located on molecular linkage group L associated with Ala were within 1.03 cM distance, indicating that they represent the single QTL from the same genomic region. Satt143 from this region explained the greatest phenotypic variation ($R^2=23.5\%$) and represents a major QTL for Ala. QTL near Satt495 is 31 cM distal to Satt143 represents a different major QTL on the same linkage group. Linkage group L also contains a major QTL

(Satt076, $R^2=24.5\%$) associated with Glu. Satt420 and Satt479 are within 4.5 cM distance on LG O, and are associated with Glu. The distance between those two molecular markers indicates that QTL for Glu is likely to be located between Satt420 and Satt479 and may be more precisely determined through fine mapping experiments. Molecular LG L had three QTL associated with Tyr within 4.35 cM distance of each other. Clearly, this indicated that LG L contains a major genomic region associated with Tyr at the same region as that controlling Ala (Table 6.5). Two QTL associated with Tyr located on LG E are within 0.5 cM distance indicating the presence of a single important genomic region containing a QTL for Tyr on that LG (Table 6.5).

An interesting aspect of QTL detected in this population for essential and non-essential amino acids is that most QTL had $R^2 \geq 10\%$, indicating that they are major QTL (Falconer and Mackay, 1996). Detection of such major QTL is useful both in MAS and positional cloning in molecular breeding programs.

The Phe, Trp and Tyr are called aromatic amino acids, and have a common biosynthesis pathway the 'Shikimate pathway', which produces the compound chorismate. Chorismate is used in the biosynthesis of Phe, Trp and Tyr (Siehl, 1999). We detected three common QTL between Phe and Tyr. There are seven different enzymes involved in the 'Shikimate

pathway' and at least three more for Phe and Tyr each, and five more for Trp. Hence it is likely that future researchers working with other populations may discover new QTL governing the biosynthesis of those enzymes.

Generally, there is not much interest in non-essential amino acids by animal feed industries or by nutritionists, despite the fact that more than half of the total ingested nitrogen comes from non-essential amino acids (Heger, 2003). QTL detected for non-essential amino acids in this study may be useful in manipulating the ratio of essential to non-essential amino acids. As mentioned before, the E:T ratio in the present population was slightly smaller than the ideal one for swine and poultry feed. Utilizing the QTL information reported in the present study will provide geneticists with opportunities to achieve the optimum ratio.

The same QTL was associated with more than one essential as well as non-essential amino acid in soybean seeds. It appears that QTL near Satt274 was the most informative QTL followed by Satt143, which was associated with nine and eight amino acids respectively (Table 6.6). Five QTL (near Satt102, Satt168, Satt196, Satt252 and Satt570) were associated with seven amino acids each.

Table 6.6: QTL† associated with various amino acids in soybean F6-derived RIL population from N87-984-16 x TN93-99.

Marker	LG	Trait
Satt274	D1b	Arg, Glu, Gly, Leu, Phe, Thr, Trp, Tyr, Val
Satt143	L	Ala, Gly, Pro, Ser, Thr, Trp, Tyr, Val
Satt102	K	Ala, Phe, Pro, Ser, Thr, Tyr, Val
Satt168	B2	Ala, Arg, Ile, Phe, Pro, Ser, Val
Satt196	K	Ala, Asp, Ile, Pro, Ser, Tyr, Val
Satt252	F	Ala, His, Ile, Leu, Ser, Tyr, Val
Satt570	G	Ala, Arg, Asp, Glu, Leu, Pro, Trp
Satt203	D1a	Asp, Glu, His, Ile, Leu, Trp
Satt292	I	Asp, Glu, Gly, Phe, Ser, Trp
Satt495	L	Ala, Gly, His, Phe, Ser, Thr
Satt177	A2	Asp, Glu, Gly, Ile, Tyr
Satt278	L	Ala, Phe, Ser, Thr, Val
Satt437	A2	Ala, Pro, Ser, Trp, Val
Satt184	D1a	His, Lys, Thr, Tyr
Satt225	A1	Ala, Arg, Tyr, Val
Satt313	L	Ala, Ile, Phe, Tyr
Satt518	K	Asp, Gly, Thr, Tyr
Satt042	A1	Ala, Gly, Val
Satt076	L	Glu, Leu, Thr
Satt133	A2	Glu, Lys, Tyr,
Satt139	C1	Gly, Ser, Trp
Satt185	E	Lys, Thr, Tyr,
Satt235	G	Ala, Asp, Lys
Satt236	A1	Leu, Thr, Trp
Satt266	D1b	Lys, Phe, Tyr
Satt268	E	Lys, Thr, Tyr
Satt420	O	Glu, Gly, Leu
Satt537	D1b	His, Trp, Val
Satt567	M	Glu, Leu, Ser
Satt600	D1b	Ala, Pro, Val
Satt002	D2	Leu, Trp
Satt172	D1b	Ala, Thr
Satt174	A1	His, Val
Satt249	J	Glu, Phe
Satt255	N	Glu, Leu
Satt380	J	His, Phe
Satt418	L	Ala, Tyr
Satt479	O	Glu, Ile
Satt512	I	His, Thr
Satt540	M	Leu, Lys
Satt564	G	His, Thr
Satt593	A1	Gly, Thr
Satt147	D1a	Asp
Satt197	B1	Leu
Satt200	A1	Leu
Satt233	A2	Ser
Satt293	H	Glu
Satt309	G	Pro
Satt427	G	Lys
Satt436	D1a	Glu
Satt458	D2	Ser
Satt461	D2	Tyr
Satt590	M	Ser

†QTL detected on the basis of default significance ($\text{LOD} \geq 2.0$) level; LG = Linkage group

Three QTL (near Satt203, Satt292 and Sattt495) located on LG D1a, I and L respectively were associated with six amino acids each (Table 6.6). Manipulation of these genomic regions may cause changes in the most amino acids concentrations. The QTL near Satt274, Satt268, Satt185 and Satt570 have been reported to be associated with nitrogen accumulation in seed during reproductive growth stages in soybean (Panthee et al., 2004), nitrogen is the major component of amino acids.

There are a few papers regarding genomic regions controlling amino acid biosynthesis in maize (*Zea mays* L) (Wang and Larkins, 2001; Wang et al., 2001a; Wang et al., 2001b) but none yet in soybean. The present paper is novel in identifying the genomic regions controlling amino acid composition and thereby providing opportunities for improvement in soybean breeding programs.

It was found that LG A1 and L had six QTL each associated with nine and 13 amino acids respectively whereas LG D1b and G had five QTL each associated with 12 and 10 amino acids respectively in this population (Table 6.7). These molecular linkage groups are informative and worth concentrating research aiming to improve the amino acid composition in soybean. Moreover, future research can be conducted to validate the molecular markers that we identified for particular amino acids throughout the genome thus confirming the QTL in future experiments.

Table 6.7: Linkage group and QTL† controlling various amino acids in a soybean F6-derived RIL population from N87-984-16 x TN93-99.

LG	Marker	Trait	No. of QTL	No. of amino acid
A1	Satt042, Satt174, Satt200, Satt225, Satt236, Satt593	Ala, Arg, Gly, His, Leu, Thr, Trp, Tyr, Val	6	9
L	Satt076, Satt143, Satt278, Satt313, Satt418, Satt495	Ala, Glu, Gly, His, Ile, Leu, Phe, Pro, Ser, Thr, Trp, Tyr, Val	6	13
D1b	Satt172, Satt266, Satt274, Satt537, Satt600	Ala, Arg, Glu, Gly, His, Leu, Lys, Phe, Pro, Thr, Trp, Tyr, Val	5	12
G	Satt235, Satt309, Satt427, Satt564, Satt570	Ala, Arg, Asp, Glu, His, Leu, Lys, Pro, Thr, Trp	5	10
A2	Satt133, Satt177, Satt233, Satt437	Ala, Asp, Glu, Gly, Ile, Lys, Pro, Ser, Trp, Tyr, , Val	4	11
D1a	Satt147, Satt184, Satt203, Satt436,	Asp, Glu, His, Ile, Leu, Lys, Thr, Trp, Tyr	4	9
D2	Satt002, Satt458, Satt461	Leu, Ser, Trp, Tyr	3	4
K	Satt102, Satt196, Satt518	Ala, Asp, Gly, Phe, Pro, Ser, Thr, Tyr, Val	3	9
M	Satt540, Satt567, Satt590	Glu, Leu, Lys, Ser	3	4
E	Satt185, Satt268	Lys, Thr, Tyr	2	3
I	Satt292, Satt512	Asp, Glu, Gly, His, Phe, Ser, Thr, Trp	2	8
J	Satt249, Satt380	Glu, His, Phe	2	3
O	Satt420, Satt479	Glu, Gly, Ile, Leu	2	4
B1	Satt197	Leu	1	1
B2	Satt168	Ala, Arg, Ile, Phe, Pro, Ser, Val	1	7
C1	Satt139	Gly, Ser, Trp	1	3
F	Satt252	Ala, His, Ile, Leu, Ser, Tyr, Val	1	7
H	Satt293	Glu	1	1
N	Satt255	Glu, Leu	1	2

†QTL detected on the basis of default significance (LOD $\geq$ 2.0) level; LG = Linkage group

The information on genomic regions controlling amino acid composition in soybean that we provide will be useful for geneticists and plant breeders to design and develop improved or better balanced amino acid composition in soybean for human and animal feeds.

References

- Baker, D.H. 2003. Ideal amino acid patterns for broiler chicks, p. 223-235, *In* J. P. F. D'Mello, ed. Amino Acids in Animal Nutrition, Second ed. CABI Publishing, Cambridge, MA.
- Boisen, S. 2003. Ideal dietary amino acid profiles for pigs, p. 157-168, *In* J. P. F. D'Mello, ed. Amino Acids in Animal Nutrition, Second ed. CABI Publishing, Cambridge, MA.
- Bourguignon, J., F. Rebeille, and R. Douce. 1999. Serine and Glycine metabolism in higher plants, p. 111-146, *In* B. K. Singh, ed. Plant amino acids: Biochemistry and biotechnology. Marcel Dekker Inc, New York.
- 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., T.E. Carter, and R.F. Wilson. 1999. Registration of Prolina soybean. *Crop Sci* 39:294-295.
- Churchill, G.A., and R.W. Doerge. 1994. Empirical threshold values for quantitative trait mapping. *Genetics* 138:963-971.

- Clarke, E.J., and J. Wiseman. 2000. Developments in plant breeding for improved nutritional quality of soybeans I. Protein and amino acids content. *J Agric Sci* 134:111-124.
- Cregan, P.B., T. Jarvik, A.L. Bush, R.C. Shoemaker, K.G. Lark, A.L. Kahler, N. Kaya, T.T. VanToai, D.G. Lohnes, and J. Chung. 1999. An integrated genetic linkage map of the soybean genome. *Crop Sci* 39:1464-1490.
- D'Mello, J.P.F. 2003. Amino acids as multifunctional molecules, p. 1-14, *In* J. P. F. D'Mello, ed. *Amino Acids in Animal Nutrition*, Second ed. CABI Publishing, Cambridge, MA.
- Falconer, D.S., and T.F.C. Mackay. 1996. *Introduction to Quantitative Genetics*. Fourth ed. Longman Group Ltd.
- Fontaine, J. 2003. Amino acid analysis of feeds, p. 15-40, *In* J. P. F. D'Mello, ed. *Amino Acids in Animal Nutrition*, Second ed. CABI Publishing, Cambridge, MA.
- Friedman, M., and D.L. Brandon. 2001. Nutritional and health benefits of soy proteins. *J Agric Food Chem* 49:1069-1086.
- Fujiwara, T., M.Y. Hirai, M. Chino, Y. Komeda, and S. Naito. 1992. Effects of sulfur nutrition on expression of the soybean seed storage protein genes in transgenic petunia. *Plant Physiol* 99:263-268.
- Heger, J. 2003. Essential to non-essential amino acid ratios, p. 103-124, *In* J. P. F. D'Mello, ed. *Amino Acids in Animal Nutrition*, Second ed. CABI Publishing, Cambridge, MA.

- Kim, H., M.Y. Hirai, H. Hayashi, M. Chino, S. Naito, and T. Fujiwara. 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.
- Lhuillier-Soundele, A., N.G. N.G. Munier-Jolain, and Ney. 1999. Influence of nitrogen availability on seed nitrogen accumulation in pea. *Crop Sci* 39:1741-1748.
- Lincoln, S.E., M.J. Daly and E.S. Lander. 1993. Constructing genetic linkage maps with MAPMAKER Version 3.: A tutorial and reference manual. Release 3rd edition. Whitehead Institute for Biomedical Research, Cambridge, MA.
- Nakasathien, S., D.W. Israel, R.F. Wilson, and P. Kwanyuen. 2000. Regulation of seed protein concentration in soybean by supra-optimal nitrogen supply. *Crop Sci* 40:1277-1284.
- Nyquist, W.E. 1991. Estimation of heritability and prediction of selection response in plant populations. *Critical Reviews in Plant Sciences* 10:235-322.
- Paek, N.C., P.J. Sexton, S.L. Laeve, and R. Shibles. 2000. Differential accumulation of soybean seed storage protein subunits in response to sulfur and nitrogen nutritional sources. *Plant prod sci* 3:268-274.
- Pantalone, V.R., F.L. Allen, and D. Landau-Ellis. 2003. Registration of 'TN93-99' Soybean germplasm. *Crop Sci* 43:1137.

- Panthee, D.R., V.R. Pantalone, C.E. Sams, A.M. Saxton, D.R. West, and W.E. Rayford. 2004. Genomic regions governing soybean seed nitrogen accumulation. *J Am Oil Chem Society* 81:77-81.
- Pazdernik, D.L., A.S. Killam, and J.H. Orf. 1997. Analysis of amino and fatty acid composition in soybean seed, using near infrared reflectance spectroscopy. *Agron J* 89:679-685.
- SAS Institute Inc. 2002. SAS guide for personal computers. 9th ed. SAS Institute, Cary, NC.
- Sexton, P.J., S.L. Naeve, N.C. Paek, and R. Shibles. 1998. Sulfur availability, cotyledon nitrogen:sulfur ratio, and relative abundance of seed storage proteins of soybean. *Crop Sci* 38:983-986.
- Siehl, D.L. 1999. The biosynthesis of Tryptophan, Tyrosine, and Phenylalanine from chorismate, p. 171-204, *In* B. K. Singh, ed. *Plant amino acids: Biochemistry and biotechnology*. Marcel Dekker Inc, New York.
- Singh, B.K. 1999. Biosynthesis of Valine, Leucine, and Isoleucine, p. 227-247, *In* B. K. Singh, ed. *Plant amino acids: Biochemistry and biotechnology*. Marcel Dekker Inc, New York.
- Soytech Inc. 2004. Oilseed statistics. Soytech Inc., Bar Harbor, ME.
- Wang, S, C, .J. Basten and Z.B. Zeng. 2003. Windows QTL Cartographer. Release V2.0. North Carolina State University, Raleigh, NC.

- Wang, X., and B.A. Larkins. 2001. Genetic analysis of amino acid accumulation in opaque-2 maize endosperm. *Plant Physiol* 125:1766-1777.
- Wang, X., D.K. Stumpf, and B.A. Larkins. 2001a. Aspartate kinase 2. A candidate gene of a quantitative trait locus influencing free amino acid content in maize endosperm. *Plant Physiol* 125:1778-1787.
- Wang, X., Y. Won, C.S. Kim, and B.A. Larkins. 2001b. Quantitative trait locus mapping of loci influencing elongation factor 1a content in maize endosperm. *Plant Physiol* 125:1271-1282.
- Zarkadas, C.G., Z. Yu, and V.D. Burrows. 1995. Protein quality of three new Canadian-developed naked oat cultivars using amino acid compositional data. *J Agric Food Chem* 43:415-421.
- Zarkadas, C.G., Z. Yu, D. Voldeng, and A. Minero Amador. 1993. Assessment of the protein quality of a new high-protein soybean cultivar by amino acid analysis. *J Agric Food Chem* 41:616-623.
- Zarkadas, C.G., H.D. Voldeng, Z.R. Yu, and V.K. Choi. 1997a. Determination of the protein quality of three new northern adapted cultivars of common and miso type soybeans by amino acid analysis. *J Agric Food Chem* 45:1161-1168.
- Zarkadas, C.G., H.D. Voldeng, Z.R. Yu, and V.K. Choi. 1999. Assessment of the protein quality of nine northern adapted yellow and brown seed coated soybean cultivars by amino acid analysis. *J Agric Food Chem* 47:5009-5018.

Zarkadas, C.G., H.D. Voldeng, Z.R. Yu, K. Shang, and P.L. Pattison.

1997b. Comparison of the protein quality of five new northern adapted natto soybean cultivars by amino acid analysis. J Agric Food Chem 45:2013-2019.

Zarkadas, C.G., Z. Yu, H.D. Voldeng, H.J. Hope, A. Minero Amador, and

J.A. Rochemont. 1994. Comparison of the protein-bound and free amino acid contents of two northern adapted soybean cultivars. J Agric Food Chem 42:21-33.

PART 7

Quantitative Trait Loci for Agronomic Traits in Soybean

This chapter is a lightly revised version of paper by the same name that will be submitted to the ***Crop Science*** by D.R. Panthee, V.R. Pantalone, A.M. Saxton, D.R. West and C.E. Sams.

My use of “our” in this chapter refers to my co-authors and myself. My primary contribution to this paper include (1) organization, development and management of all aspects of the project, (2) collecting all molecular data using polymorphic molecular markers, (3) collecting all the field data at Knoxville and Holston environments, (4) analyzing all data, and (5) all of the writing except for some minor revisions.

Abstract

Soybean [*Glycine max* (L.) Merr.] is the second most important crop after corn (*Zea mays* L.) contributing more than US\$ 15 billion annually to the US economy. There continues to be gradual improvements in seed yields of soybean by conventional breeding, but molecular techniques may provide faster genetic gains. The objective of this study was to identify quantitative trait loci (QTL) associated with those traits. To achieve this objective, 101 F₆-derived recombinant inbred lines (RIL) from a population developed from a cross of N87-984-16 x TN93-99 were used. Experiments were conducted in six total environments during 2002 and 2003. The data were analyzed using SAS software and QTL Cartographer. significant ($P < 0.05$) differences among the RIL were found

for various agronomic traits. Heritability estimates on an entry mean basis from data combined across all environments ranged from 0.12 to 0.92 for seed yield and days to flower respectively. A total of 94 polymorphic simple sequence repeat (SSR) molecular genetic markers were screened in DNA from progenies. Single factor ANOVA was used to identify candidate QTL, which were then confirmed by composite interval mapping of QTL Cartographer. Two QTL each for lodging, plant height and seed filling period, three for seed yield and six for maturity were detected in this population. We recently released the high yielding, agronomically desirable line TN03-350 as a germplasm from this population. The QTL associated with agronomic traits that we report and the recently released germplasm may be useful to improve the soybean crop in breeding programs.

Key words: agronomic traits, composite interval mapping, QTL, soybean, SSR, yield

Introduction

Soybean [*Glycine max* (L.) Merr.] is the second most important crop after corn (*Zea mays* L.) contributing more than US\$ 15 billion annually (Soytech Inc., 2004). Since the beginning of systematic soybean breeding in the 1930s in the USA, significant genetic gains have been achieved for seed yield, plant maturity and disease resistance. It is evident that there has been a continuous trend of yield improvement over time. Plant

introductions and, intra and inter-specific hybridization followed by selection were the main approaches adopted to bring about such improvements (Fehr, 1987).

With the advent of biotechnology since the 1980s, the introduction of a gene is not limited to intra or inter-specific hybridization in any crop species including soybean (Parrot and Clemente, 2004). Furthermore, efficiency of selection, even in conventional breeding, can be improved significantly with the aid of molecular approaches in terms of time and precision (Dudley, 1993; Eathington et al., 1997; Reyna and Sneller, 2001). Marker assisted selection (MAS) has already been adopted successfully in several crop species including soybean (Orf et al., 2004). Most of the traits of economic importance are quantitative in nature and hence are controlled by more than one gene. More importantly, there are some minor genes involved in this process.

Molecular markers, which are not involved in determining the trait itself but which are associated with QTL, are available in the soybean genome and have been studied extensively (Cregan et al., 1999; Shoemaker, 1994; Song et al., 2004). Such markers have been used in locating QTL and in improving quantitative traits in plant breeding via MAS. Before carrying out MAS, precise location of QTL is a prerequisite. Therefore, many researchers have located QTL of interest (Concibido et al., 2003; Lee et

al., 2001; Lee et al., 1995; Mansur et al., 1993a; Mian et al., 1998; Mian et al., 1997; Tasma and Shoemaker, 2003). Agronomic traits such as seed filling period, maturity, lodging, plant height and seed yield are quantitative traits. Seed filling period and lodging are related to seed yield, whereas plant height and lodging are correlated with each other. We were interested in mapping QTL for seed yield and traits influencing it, because yield is the most important economic trait.

Use of molecular markers in the study of quantitative traits is very important because quantitative traits are highly influenced by the growing environments but molecular markers are less influenced. The genetic basis of quantitative traits can be revealed through studies of molecular markers (Dudley, 1993). Several plant breeders have studied molecular markers in mapping QTL for use in MAS. Mansur et al. (1996) used a Minsoy x Noir 1 derived population of soybean RIL to map QTL associated with agronomic traits including plant height, lodging and yield. They found clusters of QTL associated with various agronomic traits and concluded that separation of QTL for different traits might be difficult. Since RIL have gone through several cycles of meiosis and attained homozygosity at most loci, linkages might be broken and the individual QTL for an agronomic trait could be mapped. It is important to note that detection of QTL may be population as well as environment dependent. For this reason, it is important to map QTL across multiple environments and populations to

confirm. Orf et al. (1999b) compared three populations in four environments each for QTL mapping. They found only a few common QTL across the environments and populations. This difference in detection of QTL appears to be because of differences in genetic background.

In addition to environments, there might be inter-allelic interactions making the effect of QTL complicated. Lark et al. (1995) analyzed and presented evidence of such interactions. Orf et al. (1999a) also studied the interaction of yield QTL with others. Considering all such effects, it is important to further verify the QTL controlling agronomic traits to improve the genetic gain by understanding the genetics of a trait of interest. Orf et al. (1999b) mapped QTL for yield in soybean and identified alleles from a northern adapted cultivar that could increase yield. Such alleles can be introgressed into other genotypes with MAS. The objective of our study was to identify quantitative trait loci associated with the agronomic traits of seed yield, plant lodging resistance, plant height, seed filling period, and plant maturity in soybean.

Materials and Methods

Plant materials

A total of 101 F_6 -derived recombinant inbred lines (RIL) were developed from a cross of N87-984-16 x TN93-99 and used in this study. The N87-984-16 line is one of two F_8 -derived sister lines, whose blend constitutes

the commercial cultivar 'Prolina' (Burton et al., 1999). The line TN93-99 is a high yielding and well adapted University of Tennessee breeding line and is registered as germplasm GP-280 (Pantalone et al., 2003). Crossing was made during summer of 1998. F_1 seeds were harvested in October 1998 and F_2 single plants were grown in Costa Rica during winter 1998. Generations were advanced until F_5 in Costa Rica and $F_{5:6}$ seeds were obtained in May 2000. About 300 $F_{5:6}$ single plants were grown at Knoxville as a Recombinant Inbred Line (RIL) population and 101 random plants of similar Tennessee adapted maturity were selected as a source population for this study. All 101 lines and parents were planted in 6 m length two-row plots with three replications in a randomized complete block design (RCBD) at the Knoxville Plant Science Farm (KPSF) in 2001. All the RIL, the two parents, as well as three checks ('Hutcheson', '5002T' and '5601T') were planted in an RCBD with three replications at three locations: the Knoxville and Holston units of the Knoxville Experiment Station, and the Ames Plantation Experiment Station, near Grand Junction, TN in May 2002 and 2003. Each line was planted in a four row plot of 6 m length with a spacing of 75 cm between rows.

Phenotypic traits

Days to flowering was recorded when more than 95% of the plants appeared in the full bloom (R_2) stage according to Fehr and Caviness (1977). At this stage, at least one or two flower buds on the main stem

were open. Days to flowering is the time period from date of planting to the R₂ stage.

Maturity was recorded when 95% of the pods showed their mature color, which is the R₈ stage according to Fehr and Caviness (1977). Maturity period was the number of days from date of planting to R₈. Once days to flowering and maturity were available, seed filling period was determined as their difference.

Plant height was measured as the average row height from the base of the stem at the soil surface to the top of the plant at maturity. Lodging was also scored at that time with a score range from 1 (all plants erect) to 5 (all plants prostrate). Seed yield (ton/ha) was recorded at the time of harvesting from the center two rows of every plot. Before harvesting, all plots were uniformly end-trimmed to bring the row length to 4.8 m. Seed moisture was measured at the time of harvest and yield was adjusted to 12% moisture before analyzing the data.

DNA extraction and polymerase chain reaction

DNA was extracted from the RIL and parental lines utilizing Qiagen Plant Easy DNA Extraction Kit (Qiagen Hilden, Germany). Polymerase chain reaction (PCR) consisted of 7.4 µL of ddH₂O, 1 µL of 10X PCR Buffer, 1 µL of 2 mM dNTPs mixture (Pharmacia, Piscataway, NJ), 0.5 µL of 20 µM

forward and reverse primer, 0.1 µL of 5 units/µL Klentaq (Ab Peptides Inc., St. Louis, MO) and 2 µL of 20 ng/µL template DNA. The PCR was performed in a 96-well MBS Hybaid thermocycler (Hybaid, Franklin, MA). PCR conditions were a) 94°C for five minutes followed by 35 cycles at 94°C for denaturation for 25 s, b) 47°C for annealing for 30 s, c) 72°C for 25 s for extension, and one last cycle at 72°C for final extension for five minutes. Parents were screened for a total of 585 (ATT)_n type of simple sequence repeat (SSR) genetic markers (Cregan et al., 1999), of which 138 were found polymorphic. The sequence information for the markers is publicly available from the Website (www.soybase.org), which was verified on December 16, 2004. A total of 94 out of 138 SSR markers were found to be polymorphic, and were used in QTL analysis. The remaining 44 SSR markers either did not amplify in the progeny or could not be scored because of indistinguishable levels of polymorphism.

DNA gel electrophoresis

A 6% non-denaturing polyacrylamide gel electrophoresis (PAGE) used to separate the PCR product consisted of 6% bis-acrylamide, 0.5% TBE buffer, 0.07% APS and 0.035% TEMED, which was measured as 28.5 mL of 40% bis-acrylamide, 160.2 mL of 0.593X TBE buffer, 1.33 mL of 10% APS and 66.5 µL of TEMED. Two µL of loading buffer (6X) were added to PCR products and a 10 µL sample was loaded on the gel. The running buffer was 0.5X TBE. The gel was run at constant 300 V for 3 h. A fan was

used to circulate air to keep the glass plates cool during gel running. Ethidium bromide (50 µl, 10 mg/mL) was added to the running buffer in order to visualize the bands under exposure to UV light. Bands were scored using 1 to represent P₁ (N87-984-16), 2 to represent heterozygous, and 3 to represent P₂ (TN93-99) alleles for each primer locus.

Data analysis

Phenotypic data for the agronomic traits were analyzed using the MIXED procedure to determine genotypic differences among the RIL (SAS Institute Inc, 2002). Environment and replication were considered as two blocking factors in the model. Heritability of traits was estimated by the entry mean basis (Nyquist, 1991) as follows:

$$h^2 = \frac{\sigma_g^2}{\sigma_g^2 + \left(\frac{\sigma_{ge}^2}{e} \right) + \left(\frac{\sigma^2}{re} \right)}$$

where, h^2 represents the heritability, σ_g^2 is genotypic variance, σ_{ge}^2 is genotype x environment variance, σ^2 is error variance, r is number of replications and e is the number of environments. MIXED COVTEST procedure of SAS was used to estimate REML variance components for heritability calculations.

Phenotypic correlations among agronomic traits were determined using the CORR procedure. Genetic correlations were determined using following formula (Falconer and Mackay, 1996; Kearsy and Pooni, 1996):

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

where, r_G represents genetic correlation, x represents the first trait, y represents the second trait and σ^2 is genetic variance. Cross products were generated using the MANOVA option of the GLM procedure.

Associations between molecular data and least square means of agronomic traits were detected by analyzing the data using the GLM procedure via single factor ANOVA (SAS Institute Inc, 2002). All heterozygous were excluded from the analyses because we were interested in detecting alleles which contributed additive genetic effects in our F_6 -derived RIL population. QTL position and distance were determined using Mapmaker/Exp 3.0 (Lincoln et al., 1993). Information from Mapmaker was used in QTL Cartographer (Wang et al., 2003) to confirm the QTL by composite interval mapping (CIM). We used the standard model *Zmapqtl* 6 in the CIM procedure with a 10 cM window size and 2 cM walking speed. We performed 1000 permutations on all traits in each environment to establish the empirical LOD thresholds at the 5% probability level (Churchill and Doerge, 1994). Any QTL with a LOD score

greater than the empirical LOD threshold level was declared significantly associated with the trait.

Results and Discussion

Phenotypic traits

There was a significant difference ($p < 0.05$) among the RIL for all agronomic traits: days to flower, lodging, plant height, seed filling period, maturity and seed yield in this population. Descriptive statistics of the RIL and the parents and heritability estimates for all traits are given in Table 7.1. There was transgressive segregation for all traits (Table 7.1 and Figures 7.1 to 7.3) in this population, although the magnitude was low in some of the traits. Heritability estimates varied from a low of 0.12 in seed yield to a high of 0.92 in days to flower.

Table 7.1: Descriptive statistics for agronomic traits in an F6-derived soybean population developed from N87-984-16 x TN93-99.

Trait	Minimum	Maximum	Mean	Std. Dev.	P1	P2	LSD _{0.05}	$h^2_{\ddagger}$
Days to flower	74.3	90.7	80.4	3.4	89.7	77.2	2.16	92.0
Seed filling period (days)	66.0	80.3	71.9	3.2	69.7	71.7	3.49	86.6
Maturity (days)	136.5	146.8	141.2	2.2	144.2	141.3	0.95	21.0
Lodging†	1.5	3.3	2.2	0.4	3.3	2.1	0.32	42.6
Height (cm)	59.4	98.6	82.6	7.6	96.0	74.0	3.84	62.6
Yield (t/ha)	2.3	3.7	3.0	0.2	2.7	3.2	0.25	12.2

P1 = N87-984-16; P2 = TN93-99; LSD_{0.05} = Least significant difference at the 0.05 probability level; †Lodging scored from 1 to 5 scale, 1 = All plants erect, and 5 = All plants prostrate; ‡ Heritability on an entry mean basis

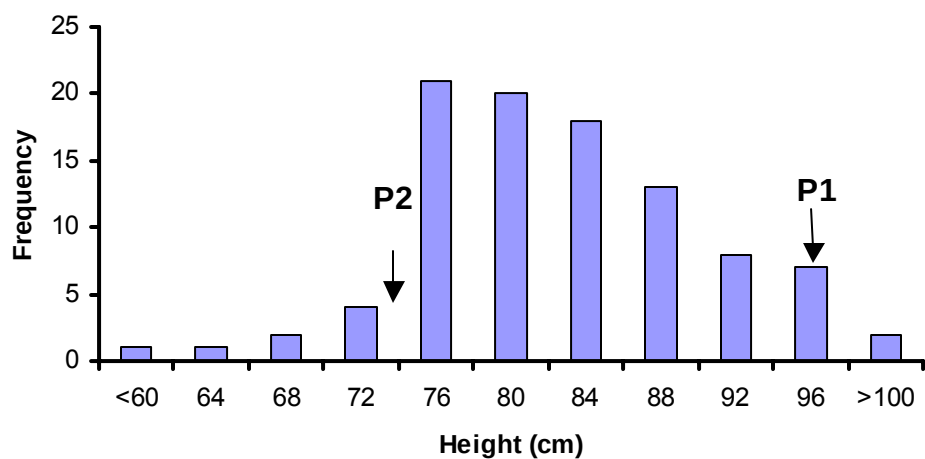


Figure 7.1: Frequency distribution for plant height in an F6-derived soybean population developed from N87-984-16 x TN93-99.

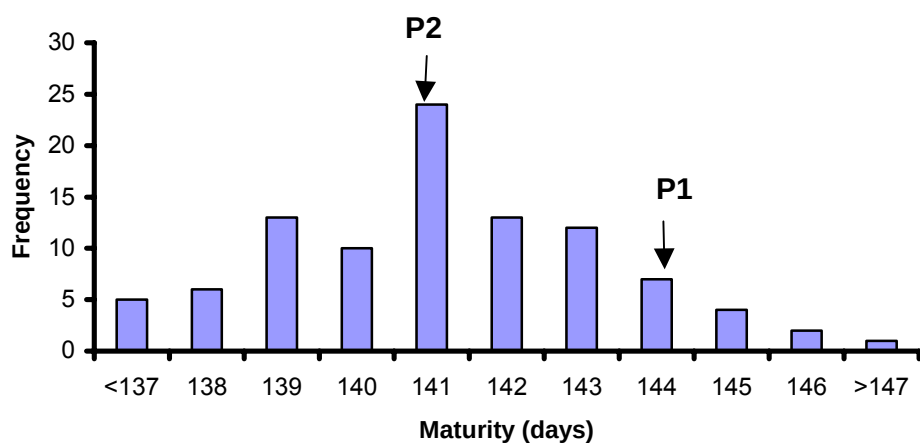


Figure 7.2: Frequency distribution for maturity in an F6-derived soybean population developed from N87-984-16 x TN93-99.

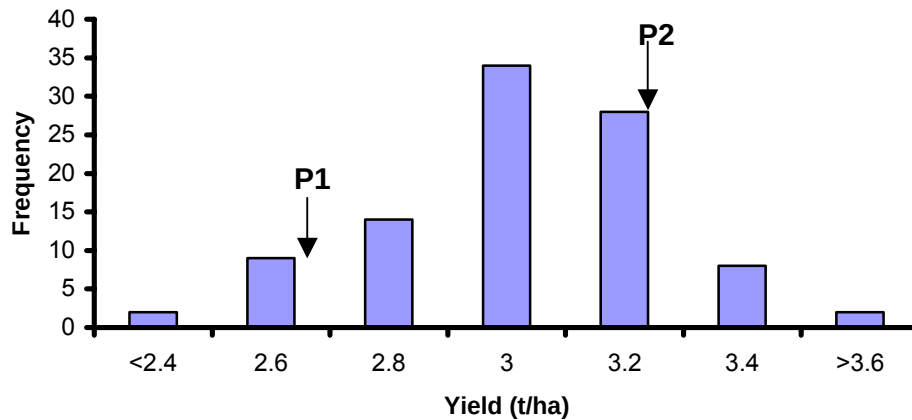


Figure 7.3: Frequency distribution for yield in an F6-derived soybean population developed from N87-984-16 x TN93-99.

Days to flower were recorded to ascertain its correlation with maturity, but only a very weak positive correlation ($r = 0.12$) was found in this population between these two traits. This indicated that early flowering lines do not necessarily mature first, which means those lines keep on accumulating photosynthate in the seed and can achieve high yield. This is apparent from the strong negative correlation between days to flower and seed filling period ($r = -0.60$) (Table 7.2). longer seed filling period allows greater accumulation of photosynthate in the seed, hence higher yield. This is supported by the positive correlation between seed filling period and seed yield ($r = 0.41$). It is interesting to note that there was a positive correlation between days to flower and lodging ($r = 0.58$). This

Table 7.2: Genetic and phenotypic correlation coefficients among agronomic traits in an F6-derived soybean population developed from N87-984-16 x TN93-99.

Trait	Seed filling period	Maturity	Lodging	Height	Yield
Days to flower	-0.60***	0.12*	0.58***	0.42***	-0.39***
Seed filling period	-	0.72***	-0.28***	-0.02 ^{ns}	0.41***
Maturity	NA	-	0.17**	0.35***	0.17**
Lodging	NA	0.03 ^{ns}	-	0.60***	-0.35***
Height	NA	0.49***	0.58***	-	-0.29***
Yield	NA	0.06 ^{ns}	-0.94***	-0.11 ^{ns}	-

Values below the diagonal are genetic correlations and those above the diagonal are phenotypic correlations; NA = Not available since data for days to flower were recorded in only one location; *, **, *** = Significant at 0.05, 0.01 and 0.001 probability levels respectively; ns = Non-significant

might be because nutrients are diverted towards vegetative parts making them tall and succulent (because of enlarged cell size) making the plants prone to lodging. On the other hand, early flowering plants divert energy towards the reproductive system which reduces vegetative growth, hence they are not succulent, resulting in reduced lodging. This explanation is further supported by a positive correlation between plant height and lodging ($r = 0.60$) in this population (Table 7.2). Similar observations have been reported (Mansur et al., 1996). Lodging caused reduced yield, which is apparent from the negative correlation between lodging and yield ($r = -0.35$) in this population.

Unlike other studies (Kabelka et al., 2004; Lee et al., 1996; Mansur et al., 1993b; Orf et al., 1999b), there was low heritability for the agronomic traits, especially seed yield in the present population. However, our

observations were within the range reported in other studies (Alliprandini and Vello, 2004; Burton, 1987).

QTL analysis

There were four QTL linked to molecular markers (Satt139, Satt524, Satt518 and Satt418) for days to flower detected by single factor ANOVA, however we could not find QTL when data were analyzed by CIM of QTL Cartographer. There were two QTL linked to molecular markers (Satt042 and Satt570) located on LG A1 and G respectively associated with seed filling period in this population of soybean (Table 7.3 and Figure 7.4). Considering the coefficient of determination (R^2) value, QTL near Satt042 is a major QTL for seed filling period (Table 7.3). Phenotypic variation explained by an individual seed filling period QTL ranged from 15.6 to 22.7% in this population. There was a small difference of two days for seed filling period between the parents. Probably because of this limited difference, allelic contributions from both the parents were present in this population. There are five QTL for seed filling period reported in Soybase (Keim et al., 1990; Mansur et al., 1993b) but none of them are from linkage group A1 or G. Therefore the two QTL near Satt042 and Satt570 detected in the present population are novel.

There were two QTL (near Satt184 and Satt292) associated with plant height located on linkage groups D1a and I respectively (Table 7.3). N87-

Table 7.3: Quantitative trait loci, their position and genetic contribution as additive effects for agronomic traits in an F6-derived soybean population developed from N87-984-16 x TN93-99.

Trait	Environment	LG	QTL	Position (cM)	LOD Score	Additive R ² (%) Effect#	
Seed filling	KPSF 2003	A1	Satt042	27.66	3.8	-2.28	22.7
	KPSF 2002	G	Satt570	12.73	3.0	1.57	15.6
Plant height	Holston 2002	D1a	Satt184	17.52	2.9	2.52	10.2
	Holston 2002	I	Satt292	82.78	3.0	3.10	14.4
Lodging	Combined 2003	A1	Satt225	95.16	3.9	-3.00	20.6
	Combined (2002-03)	A1	Satt593	25.56	5.2	0.24	32.7
Maturity	Ames 2003	A1	Satt042	27.66	2.9	-1.01	12.4
	KPSF 2003	A1	Satt042	27.66	3.5	-1.39	13.5
	Combined (2002-03)	A1	Satt591	31.14	3.5	-0.95	12.8
	Combined 2003	A1	Satt591	31.14	3.0	-0.95	10.9
	Combined 2002	E	Satt263	45.40	3.8	0.93	13.8
	Ames 2002	F	Satt252	16.08	3.0	0.78	11.7
	Ames 2003	H	Satt293	89.09	3.5	1.17	18.0
	Combined 2003	H	Satt293	89.09	3.4	1.09	15.3
	Combined 2002	I	Satt292	82.78	3.2	1.27	25.7
Yield	KPSF 2002	A1	Satt042	27.66	3.3	-0.93	14.6
	Holston 2002	D1b	Satt412	72.57	3.2	-0.07	13.3
	Combined 2003	L	Satt076	61.35	2.9	-1.25	13.8
	Holston 2003	L	Satt076	61.35	4.1	-2.01	21.0

LG = Linkage group; QTL = Quantitative trait loci; KPSF = Knoxville Plant Science Farm, # ½(P1-P2)

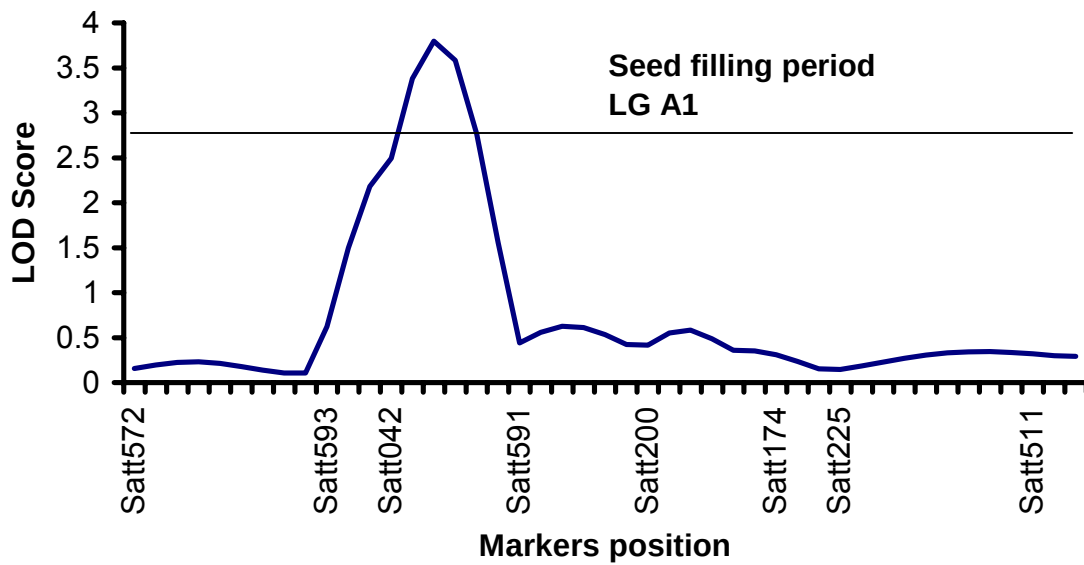


Figure 7.4: LOD score plot for seed filling period on LG A1 in an F6-derived soybean population developed from N87-984-16 x TN93-99.

984-16 is taller than TN93-99 (Table 7.1). Both the loci received the high alleles through N87-984-16 in this population. Phenotypic variation for plant height explained by an individual QTL ranged from 10.2 to 14.4%. At least one plant height QTL near A295_1 from linkage group D1a (Lark et al., 1995) has been reported but this is more than 70 cM from Satt184, indicating that they may be independent. Hence QTL near Satt184 detected in the present population marks a novel QTL associated with plant height in soybean. Two QTL near Satt127 and Satt440 associated with plant height from linkage group I have been reported (Sebolt et al., 2000; Yuan et al., 2002). Satt127 is about 47 cM upstream of Satt292

whereas Satt440 is about 30 cM downstream. Both distances are not very close to Satt292 indicating that, most probably, QTL near Satt292 is also a new QTL associated with plant height.

There were two QTL linked to molecular markers Satt225 and Satt593 associated with lodging and located on linkage group A1 (Table 7.3 and Figure 7.5). The QTL near Satt593 received the increased allele from N87-984-16 whereas the one near Satt225 received the increased allele from TN93-99. The TN93-99 line is better than N87-984-16 in lodging resistance (Table 7.1). However, since allelic contribution was found from both parents, the alleles for improvement are likely to be dispersed in both the parents. Phenotypic variation for lodging explained by an individual QTL ranged from 20.6% for Satt225 to 32.7% for Satt593 loci. At least 34 QTL associated with lodging have already been reported in soybase (www.soybase.org) by various researchers. Out of those, at least three QTL linked to markers A082_1, A104_1 and A170_1 are from linkage group A1. Satt225 detected in the present population is within 3 cM downstream of A104_1 and A170_1 whereas it is about 7.1 cM upstream of A082_1 in composite integrated linkage map of soybean (Song et al., 2004) indicating that Satt225 is most likely in the same genomic region of an already reported QTL. However, QTL near Satt593 is novel since no QTL have been reported in close proximity.

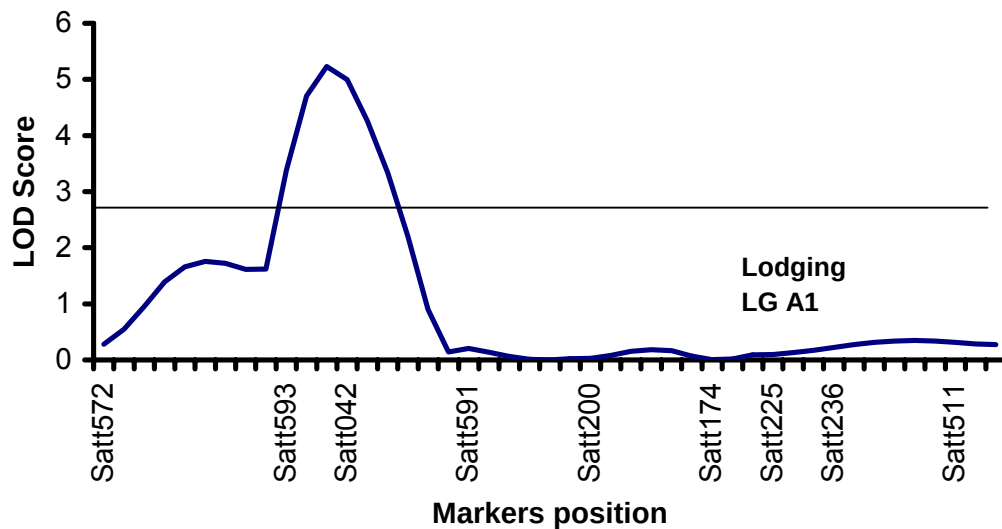


Figure 7.5: LOD score plot for lodging on LG A1 in an F6-derived soybean population developed from N87-984-16 x TN93-99.

Six QTL near Satt042, Satt252, Satt263, Satt292, Satt293 and Satt591 were associated with maturity in this population. They were distributed in five molecular linkage groups A1, E, F, H and I (Table 7.3 and Figure 7.6).

Two QTL were detected from LG A1 indicating that this is an important linkage group controlling maturity in southern adapted soybean. Satt042 and Satt591 are within 5 cM suggesting that this region has a single QTL controlling maturity on this linkage group (Figure 7.6). QTL near Satt292 located on LG I was detected as a major QTL, which explained 25.7% of the phenotypic variation and contributed the largest additive effect. The least phenotypic variation (10.9%) was explained by Satt591 located on

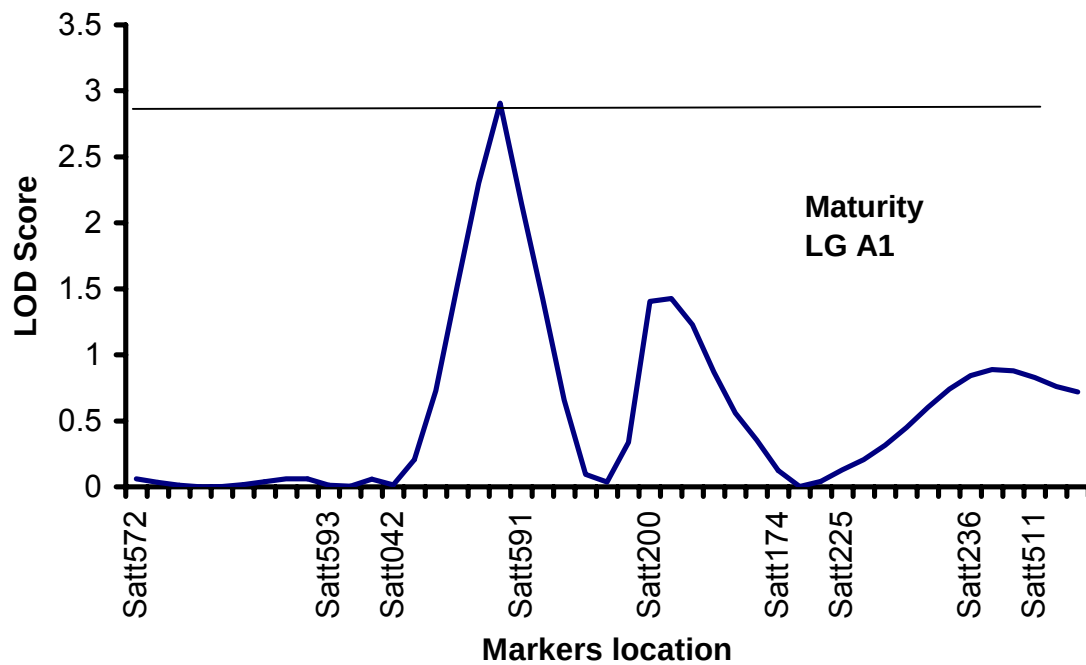


Figure 7.6: LOD score plot for maturity on LG A1 in an F6-derived soybean population developed from N87-984-16 x TN93-99.

LG A1. High alleles in four of six QTL were contributed through N87-984-16, which was the later maturing parent (Table 7.1). A maturity associated QTL (near Satt050 located on LG A1) was previously detected in a population of Essex x William (Chapman et al., 2003) and is 15.3 cM downstream of Satt591, which we detected in the present population. Moreover, QTL near Satt382 was detected in another population recently (Kabelka et al., 2004) and is within 1.3 cM upstream of Satt042 and within 4.7 cM upstream of Satt591. Two QTL associated with maturity that we detected on LG A1 in the same genomic region, further indicating that this is an important region controlling maturity in soybean.

There is one maturity QTL near BLT030_1 on linkage group F (Specht et al., 2001), which is about 7.5 cM upstream of Satt252, which we detected as a maturity QTL in our population suggesting that both the markers reveal a single QTL in this region. There are at least two maturity QTL near A688_1 and Satt127 reported from linkage group I (Sebolt et al., 2000) but both of them are at least 47.4 cM upstream indicating that QTL near Satt292 is most likely an independent QTL on that LG.

Since there are no QTL associated with maturity from molecular linkage group E or H reported previously, QTL near Satt263 and Satt293 detected in our population are new. QTL near Satt292 is also a new QTL since other reported QTL from LG I are not in the same genomic region but Satt042 and Satt591 are not new rather verify existing QTL since they are in the same genomic regions as those reported before. The present experiment verifies that LG A1 has an important genomic region around Satt042 and Satt591 controlling maturity in soybean, and provides new information on additional QTL controlling soybean maturity.

There were three QTL near Satt042, Satt412 and Satt076 associated with seed yield in this population. Those QTL were distributed on three molecular linkage groups A1, D1b and L respectively (Table 7.3; Figures 7.7 and 7.8). Phenotypic variation explained by a single QTL for seed yield

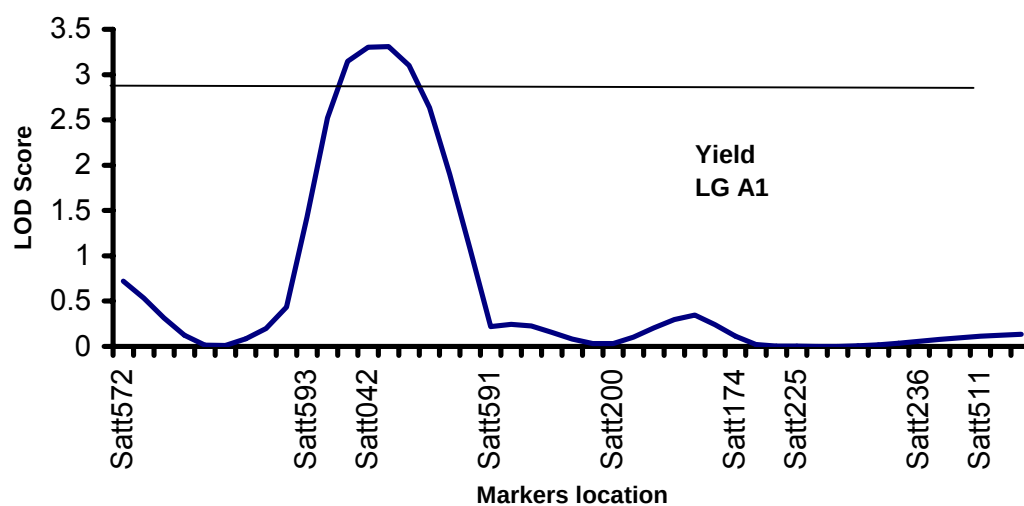


Figure 7.7: LOD score plot for seed yield on LG A1 in an F6-derived soybean population developed from N87-984-16 x TN93-99.

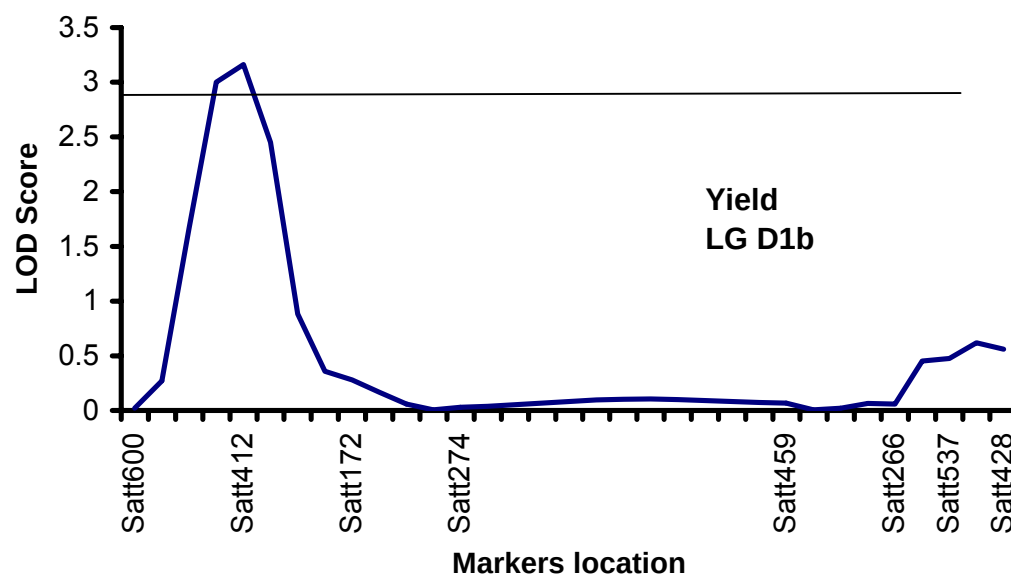


Figure 7.8: LOD score plot for seed yield on LG D1b in an F6-derived soybean population developed from N87-984-16 x TN93-99.

ranged from 13.3 to 21.1%. Interestingly, all the high alleles were contributed by TN93-99, which was the high parent for seed yield (Table 7.1). QTL near Satt076 had the greatest additive genetic effect followed by Satt042, which explained the most phenotypic variation. These three QTL explained about 41% of the total phenotypic variation for seed yield in this population.

Recently, 15 QTL associated with seed yield were detected (Kabelka et al., 2004) but only one QTL was in the same genomic region that we detected in our population. Kabelka et al. (2004) found a QTL near Satt382 associated with yield, which is within 1.3 cM upstream of Satt042 detected in our population. There are at least 29 QTL associated with yield already reported in soybase (www.soybase.org). One of those is Satt561, located on LG L (Orf et al., 1999a) and it is within 10 cM downstream of the Satt076 detected in our population. Both the QTL near Satt042 and Satt076 are within the genomic region of a QTL already reported. However, QTL near Satt412 (LG D1b) is novel, and there are no yield QTL reported previously from linkage group D1b.

A total of 12 QTL were found to be associated with at least one of the agronomic traits in this population (Table 7.4). The QTL near Satt042 was associated with three agronomic traits seed filling period, maturity and yield whereas the one near Satt292 was associated with two agronomic

Table 7.4: Number of QTL in various linkage groups associated with agronomic traits in an F6-derived soybean population developed from N87-984-16 x TN93-99.

Linkage Group	Agronomic traits					Total QTL
	Lodging	Plant height	Seed filling	Maturity	Yield	
A1	Satt225, Satt593	-	Satt042	Satt042, Satt591	Satt042	4
D1a	-	Satt184	-	-	-	1
D1b	-	-	-	-	Satt412	1
E	-	-	-	Satt263	-	1
F	-	-	-	Satt252	-	1
G	-	-	Satt570	-	-	1
H	-	-	-	Satt293	-	1
I	-	Satt292	-	Satt292	-	1
L	-	-	-	-	Satt076	1
Total QTL	2	2	2	6	3	12

trait plant height and maturity. All the remaining QTL were associated with only one agronomic trait each.

Molecular linkage group A1 contained four QTL associated with four traits and LG I contained one QTL associated with two traits. The QTL near Satt042 located on LG A1 was associated with seed filling period, maturity and seed yield. A single gene controlling more than one trait, termed pleiotrophy, is helpful to improve more than one trait at a time if the genetic effects are in the desired direction.

Interestingly, QTL near Satt042 had the additive genetic effect in the same direction for all three traits in this population (Table 7.3). So was the case

with QTL near Satt292 with plant height and maturity. The remaining seven linkage groups contained only one QTL associated with one trait each (Table 7.4) indicating that most of the QTL associated with multiple agronomic traits were clustered on linkage group A1.

Identifying the favorable alleles at important loci and introducing them into the genetic background of a desirable genotype is a common practice to improve the agronomic performance of traits of interest (Smalley et al., 2004). Generally desirable alleles are contributed through the high parent but this is not true in all cases for a quantitative trait. In the present experiment, we found the favorable contributing alleles from both the parents, which is common for quantitative traits (Csanadi et al., 2001; Mian et al., 1997). Once favorable alleles are identified, it is easier to introgress such alleles into the background of desirable genotype (Concibido et al., 2003).

We were able to develop a high yielding line from this population, which yielded 3.7 t/ha, significantly higher than the commercial check cultivar 5601T. Our new line was released as germplasm TN03-350 recently. It is anticipated that this germplasm will serve as a valuable parent in future breeding programs to improve seed yield in soybean.

References

- Alliprandini, L.F., and N.A. Vello. 2004. Heritability and correlations among traits in four-way soybean crosses. *Euphytica* 136:81-91.
- Burton, J.W. 1987. Quantitative genetics: Results relevant to soybean breeding, p. 211-247, *In* J. R. Wilcox, ed. Soybeans: Improvement, Production and Uses, Second ed. American Society of Agronomy, Inc., Madison, WI, USA.
- Burton, J.W., T.E. Carter, and R.F. Wilson. 1999. Registration of Prolina soybean. *Crop Sci* 39:294-295.
- Chapman, A., V.R. Pantalone, A. Ustun, F.L. Allen, D. Landau Ellis, R.N. Trigiano, and P.M. Gresshoff. 2003. Quantitative trait loci for agronomic and seed quality traits in an F2 and F4:6 soybean population. *Euphytica* 129:387-393.
- Churchill, G.A., and R.W. Doerge. 1994. Empirical threshold values for quantitative trait mapping. *Genetics* 138:963-971.
- Concibido, V.C., B. La Vallee, P. McLaird, N. Pineda, J. Meyer, L. Hummel, J. Yang, K. Wu, and X. Delannay. 2003. Introgression of a quantitative trait locus for yield from Glycine soja into commercial soybean cultivars. *Theor Appl Genet* 106:575-582.
- Cregan, P.B., T. Jarvik, A.L. Bush, R.C. Shoemaker, K.G. Lark, A.L. Kahler, N. Kaya, T.T. VanToai, D.G. Lohnes, and J. Chung. 1999. An integrated genetic linkage map of the soybean genome. *Crop Sci* 39:1464-1490.

- Csanadi, G., J. Vollmann, G. Stift, and T. Lelley. 2001. Seed quality QTLs identified in a molecular map of early maturing soybean. *Theor Appl Genet* 103:912-919.
- Dudley, J.W. 1993. Molecular markers in plant improvements: Manipulations of genes affecting quantitative traits. *Crop Sci* 33:660-668.
- Eathington, S.R., J.W. Dudley, and G.K.R. II. 1997. Usefulness of marker-QTL associations in early generation selection. *Crop Sci* 37:1686-1693.
- Falconer, D.S., and T.F.C. Mackay. 1996. *Introduction to Quantitative Genetics*. Fourth ed. Longman Group Ltd.
- Fehr, W.R. 1987. Breeding methods for cultivar development, p. 249-293, *In* J. R. Wilcox, ed. *Soybeans: Improvement, production and uses*, Vol. 16, Second ed. Crop Science Society of America, Madison, WI, USA.
- Fehr, W.R., and C.E. Caviness. 1977. *Stages of soybean development*. Iowa State University, Ames, IA.
- Kabelka, E.A., B.W. Diers, W.R. Fehr, A.R. LeRoy, I.C. Baianu, T. You, D.J. Neece, and R.L. Nelson. 2004. Putative alleles for increased yield from soybean Plant Introduction. *Crop Sci* 44:784-791.
- Kearsey, M.J., and H.S. Pooni. 1996. *The Genetical Analysis of Quantitative Traits*. First ed. Chapman and Hall, New York.

- Keim, P., B.W. Diers, T.C. Olson, and R.C. Shoemaker. 1990. RFLP mapping in soybean: association between marker loci and variation in quantitative traits. *Genetics* 126:735-742.
- Lark, K.G., K. Chase, F. Adler, L.M. Mansur, and J.H. Orf. 1995. Interactions between quantitative trait loci in soybean in which trait variation at one locus is conditional upon a specific allele at another. *Proc Natl Acad Sci U S A* 92:4656-4660.
- Lee, S.H., K.Y. Park, H.S. Lee, E.H. Park, and H.R. Boerma. 2001. Genetic mapping of QTLs conditioning soybean sprout yield and quality. *Theor Appl Genet* 103:702-709.
- Lee, S.H., M.A. Bailey, M.A.R. Mian, E.R. Shipe, D.A. Ashley, W.A. Parrott, R.S. Hussey, and H.R. Boerma. 1996. Identification of quantitative trait loci for plant height, lodging, and maturity in a soybean population segregating for growth habit. *Theor Appl Genet* 92:516-523.
- Lee, S.H., J.C. Kim, M.S. Lee, W.D. Heo, H.Y. Seo, H.W. Yoon, J.C. Hong, S.Y. Lee, J.D. Bahk, and I. Hwang. 1995. Identification of a novel divergent calmodulin isoform from soybean, which has differential ability to activate calmodulin-dependent enzymes. *J Biol Chem* 270:21806-21812.
- Lincoln, S.E., M.J. Daly and E.S. Lander. 1993. Constructing genetic linkage maps with MAPMAKER Version 3.: A tutorial and reference

manual. Release 3rd edition. Whitehead Institute for Biomedical Research, Cambridge, MA.

Mansur, L.M., J. Orf, and K.G. Lark. 1993a. Determining the linkage of quantitative trait loci to RFLP markers using extreme phenotypes of recombinant inbreds of soybean (*Glycine max* L. Merr.). *Theor Appl Genet* 86:914-918.

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

Mansur, L.M., J.H. Orf, K. Chase, T. Jarvik, P.B. Cregan, and K.G. Lark. 1996. Genetic mapping of agronomic traits using recombinant inbred lines of soybean. *Crop Sci* 36:1327-1336.

Mian, M.A.R., D.A. Ashley, W.K. Vencill, and H.R. Boerma. 1998. QTLs conditioning early growth in a soybean population segregating for growth habit. *Theor Appl Genet* 97:1210-1216.

Mian, M.A.R., E.R. Shipe, J. Alvernaz, J.D. Mueller, D.A. Ashley, and H.R. Boerma. 1997. QTLs associated with chlorimuron ethyl sensitivity in soybean: effects on seed yield and related traits. *Theor Appl Genet* 94:971-974.

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

- Orf, J.H., B.W. Diers, and H.R. Boerma. 2004. Genetic Improvement: Conventional and Molecular-Based Strategies, p. 417-450, *In* H. R. Boerma and J. E. Specht, eds. SOYBEANS: Improvement, Production, and Uses, Third ed. American Society of Agronomy, Madison, WI, USA.
- Orf, J.H., K. Chase, F.R. Adler, L.M. Mansur, and K.G. Lark. 1999a. Genetics of soybean agronomic traits. II. Interactions between yield quantitative trait loci in soybean. *Crop Sci* 39:1652-1657.
- Orf, J.H., K. Chase, T. Jarvik, L.M. Mansur, P.B. Cregan, F.R. Adler, and K.G. Lark. 1999b. Genetics of soybean agronomic traits. I. Comparison of three related recombinant inbred populations. *Crop Sci* 39:1642-1651.
- Pantalone, V.R., F.L. Allen, and D. Landau-Ellis. 2003. Registration of 'TN93-99' Soybean germplasm. *Crop Sci* 43:1137.
- Parrot, W.A., and T.E. Clemente. 2004. Transgenic soybean, p. 265-302, *In* H. R. Boerma and J. E. Specht, eds. Soybeans: Improvement, production and uses. Crop Science Society of America, Madison, WI, USA.
- Reyna, N., and C.H. Sneller. 2001. Evaluation of marker-assisted introgression of yield QTL alleles into adapted soybean. *Crop Sci* 41:1317-1321.
- SAS Institute Inc. 2002. SAS guide for personal computers. 9th Edition ed. SAS Institute, Cary, NC.

- Sebolt, A.M., R.C. Shoemaker, and B.W. Diers. 2000. Analysis of a quantitative trait locus allele from wild soybean that increases seed protein concentration in soybean. *Crop Sci* 40:1438-1444.
- Shoemaker, B.C. 1994. RFLP map of soybean. *Adv cell mol biol plants* 1:299-309.
- Smalley, M.D., W.R. Fehr, S.R. Cianzio, F. Han, S.A. Sebastian, and L.G. Streit. 2004. Quantitative trait loci for soybean seed yield in elite and plant introduction germplasm. *Crop Sci* 44:436-442.
- Song, Q.J., L.F. Marek, R.C. Shoemaker, K.G. Lark, V.C. Concibido, X. Delannay, J.E. Specht, and P.B. Cregan. 2004. A new integrated genetic linkage map of the soybean. *Theor Appl Genet* 109:122 - 128.
- Soytech Inc. 2004. Oilseed statistics. Soytech Inc., Bar Harbor, ME.
- Specht, J.E., K. Chase, M. Macrander, G.L. Graef, J. Chung, J.P. Markwell, M. Germann, J.H. Orf, and K.G. Lark. 2001. Soybean response to water: A QTL analysis of drought tolerance. *Crop Sci* 41:493-509.
- Tasma, I.M., and R.C. Shoemaker. 2003. Mapping flowering time gene homologs in soybean and their association with maturity (E) loci. *Crop Sci* 43:319-328.
- Wang, S, C, .J. Basten and Z.B. Zeng. 2003. Windows QTL Cartographer. Release V2.0. North Carolina State University, Raleigh, NC.

Yuan, J., V.N. Njiti, K. Meksem, M.J. Iqbal, K. Triwitayakorn, M.A. Kassem, G.T. Davis, M.E. Schmidt, and D.A. Lightfoot. 2002. Quantitative trait loci in two soybean recombinant inbred line populations segregating for yield and disease resistance. *Crop Sci* 42:271-277.

PART 8

**Registration of Improved Soybean Protein Germplasms TN03-350,
TN04-5321 and TN04-5363**

This chapter is a lightly revised version of paper by the same name that has been submitted to the ***Crop Science*** in 2004 by Dilip R. Panthee and Vincent R. Pantalone.

My use of “our” in this chapter refers to my co-authors and myself. My primary contribution to this paper include (1) organization, development and management of all aspects of the project, (2) collecting all molecular data using polymorphic molecular markers, (3) collecting all the field data at Knoxville and Holston environments, (4) analyzing for protein and amino acids composition, (5) analyzing all data, and (6) all of the writing except for some minor revisions.

The soybean [*Glycine max* (L.) Merr.] germplasms lines TN03-350 (Reg. no. GP-), TN04-5321 (Reg. no. GP-) and TN04-5363 (Reg. no. GP-.....) were developed by the Tennessee Agricultural Experiment Station and released in April 2004. TN03-350 and TN04-5321 lines combine high yield and increased protein concentrations whereas TN04-5363 combines high yield and increased sulfur containing amino acids cysteine and methionine. All three lines TN03-350, TN04-5321 and TN04-5363 are F₆-derived lines from the cross N87-984-16 x TN93-99. The N87-984-16 parent is one of the two F₈-derived sister lines of N87-984, a blend of which constitutes the high protein commercial

cultivar 'Prolina' (Burton et al., 1999). Prolina was selected from the first cycle of a recurrent selection population designated NRS5. The population was originated from the matings of 10 high protein lines with the cultivars Bragg (Hinson and Hartwig, 1964), Ransom (Brim and Elledge, 1973) and Davis (Caviness and Walters, 1966). Average protein and oil concentration in Prolina are 461 g kg⁻¹ and 200 g kg⁻¹ seed respectively (Burton et al., 1999). TN93-99 is an F₄-derived high yielding germplasm developed from a triparental crossing (TN85-55 x TN5-85) x Hutcheson (Pantalone et al., 2003a). TN93-99 is the registered germplasm GP280 (<http://www.ars-grin.gov>).

Crossing of N87-984-16 x TN93-99 was made during the summer of 1998. The F₁ seeds were harvested in October 1998 and F₁ plants were grown in Costa Rica during winter 1998-99. Generations were advanced via modified single seed descent until F₅ in Costa Rica, and F_{5:6} seeds were obtained in May 2000. About 300 F_{5:6} single plants were grown at Knoxville, TN as a recombinant inbred line (RIL) population, and 101 random plants of similar maturity were selected as a source population for evaluation. All 101 RIL and parents were planted in two-row plots of 6.1m length with three replications in a randomized complete block design (RCBD) at the Plant Sciences Unit, Knoxville Experiment Station, Knoxville, TN in 2001. All RIL, parents and two checks '5002T' (Pantalone

et al., 2004) and '5601T' (Pantalone et al., 2003b) were grown in four-row plots of 6.1m length at three locations (the Holston and Plant Sciences Units of the Knoxville Experiment Station, and Ames Plantation, Grand Jackson, TN) with three replications in 2002 and 2003. On the basis of strong field performance in 2002, one of the $F_{6:8}$ RIL was designated as TN03-350 and was included in the 2003 MG V USDA Southern Regional Uniform Preliminary Test. In that test, agronomic, protein, oil and disease reaction data were collected from 11 locations.

Line TN03-350 was released as a germplasm because of its high yield potential in TN and increased protein concentration throughout broad geographical regions of southern USA. TN03-350 was the highest protein containing line (439 g kg^{-1}) out of 48 entries included in the 2003 MG V USDA Southern Regional Uniform Preliminary Test in 11 locations (Paris, 2003). However, yield of TN03-350 was significantly lower than both 5601T and 5002T averaged over 11 locations in USDA Southern Regional Uniform Preliminary Test of 2003. TN03-350 outperformed the parents and checks for seed yield and protein concentrations in six TN environments. Average protein concentration of TN03-350 (426 g kg^{-1}) was significantly higher than that of the checks 5601T (404 g kg^{-1}) and 5002T (391 g kg^{-1}) averaged over six TN environments (2002-2003). It has an average level (197 g kg^{-1} seed) of oil content. The seed yield of TN03-350 at 3655 kg ha^{-1} was significantly higher than that of the commercial

check cultivar 5601T at 3347 kg ha⁻¹ and 5002T at 3272 kg ha⁻¹ averaged over a total of six testing environments in Tennessee. The information obtained from these evaluations indicates that TN03-350 may be useful for future soybean research and development.

TN03-350 has white flower color, gray pubescence, tan pod wall, and a determinate growth habit. TN03-350 is resistant to stem canker [caused by *Diaporthe phaseolorum* (Cooke and Ellis) Sacc. Var. *caulivora* K.L. Athow & R.M. Caldwell]. TN03-350 matures about four days later than 5601T, thus its relative maturity is approximately 6.0. Seeds are yellow with smooth seed coats and buff hila. The seed size is approximately 15 g 100 seeds⁻¹ with excellent seed quality.

TN04-5321 was released as a germplasm because of its high protein concentration. TN04-5321 outperformed the parents and checks for seed protein concentration. It outperformed the parents and was on par with commercial checks for seed yield averaged over six TN environments (2002-2003). Average protein concentration of TN04-5321 (431 g kg⁻¹ seed) was significantly higher than that of the checks 5601T (405 g kg⁻¹ seed) and 5002T (392 g kg⁻¹ seed) as well as the parents N87-984-16 (427 g kg⁻¹ seed) and TN93-99 (390 g kg⁻¹ seed). The seed yield of TN04-5321 at 3222 kg ha⁻¹ was not significantly different from the commercial

check cultivars 5601T at 3347 kg ha⁻¹ and 5002T at 3272 kg ha⁻¹, averaged over a total of six testing environments in Tennessee.

TN04-5321 has white flower color, gray pubescence, tan pod wall, and a determinate growth habit. TN04-5321 is similar to commercial checks for lodging. TN04-5321 matures about three days later than 5601T, thus its relative maturity is approximately 5.9. The seeds are yellow with smooth seed coats and buff hila. The seed size is large, exceeding 16 g 100 seeds⁻¹ with excellent seed quality.

TN04-5363 was released as a germplasm because of its high total sulfur containing amino acids and high yield. TN04-5363 was significantly higher than the parents and checks for cysteine concentration, which contributed to its high total sulfur containing amino acids. Average total sulfur containing amino acid concentration (33 g kg⁻¹ protein) is very close to the World Health Organization standard for animal feed of 35 g kg⁻¹ protein, which is based on egg protein (George and de Lumen, 1991). TN04-5363 has a normal level of seed protein (399 g kg⁻¹) and oil (208 g kg⁻¹) content. TN04-5363 exhibited a comparable level of seed yield as the check cultivars. The seed yield of TN04-5363 at 3231 kg ha⁻¹ was not significantly different from that of the commercial check cultivars 5601T at 3347 kg ha⁻¹ and 5002T at 3272 kg ha⁻¹ averaged over a total of six testing environments in Tennessee. TN04-5363 matures 4 days earlier

than 5601T and 2 days earlier than 5002T, thus its relative maturity is 4.9. Interestingly, despite the negative correlation between maturity and seed yield exhibited in the population from which this line was derived, TN04-5363 has maintained high yield with earlier maturity. TN04-5363 may be useful as a parent to improve total sulfur containing amino acid, seed yield and provide maturity that is well adapted to the Mid-South region.

TN04-5363 has purple flower color, gray pubescence, tan pod wall, and a determinate growth habit. TN04-5363 is similar to commercial checks for lodging. The seeds are yellow in color with smooth seed coats and buff hila. The seed size is large, approximately 16 g 100 seeds⁻¹ with excellent seed quality.

Breeder seed stock of TN03-350, TN04-5321 and TN04-5363 will be maintained by the Tennessee Agricultural Experiment Station for five years in cold storage, and samples of 200-300 seeds will be distributed to breeders and researchers upon request.

References

- Brim, C.A. and C. Elledge. 1973. Registration of 'Ransom' soybeans. Crop Sci. 13:130.
- Burton, J.W., T.E. Carter and R.F. Wilson. 1999. Registration of Prolina soybean. Crop Sci 39:294-295.

- Caviness, C.E. and H.J. Walters. 1966. Registration of 'Davis' soybeans. Crop Sci. 6:502.
- George, A.A. and B.O. de Lumen. 1991. A novel methionine-rich protein in soybean seed: identification, amino acid composition, and N-terminal sequence. J Agric Food Chem 39:224-227.
- Hinson, K. and E.E. Hartwig. 1964. 'Bragg' and 'Hardee' soybeans. Crop Sci. 4:664.
- Pantalone, V.R., F.L. Allen and D. Landau-Ellis. 2003a. Registration of 'TN93-99' Soybean germplasm. Crop Sci 43:1137.
- Pantalone, V.R., F.L. Allen and D. Landau-Ellis. 2003b. Registration of '5601T' Soybean. Crop Sci. 43:1123-1124.
- Pantalone, V.R., F.L. Allen and D. Landau-Ellis. 2004. Registration of '5002T' Soybean. Crop Sci. 44:1483-1484.
- Paris, B. 2003. Uniform Soybean Test: Southern States 2003. USDA-ARS, Stoneville, MS.

PART 9

CONCLUSION AND FUTURE RESEARCH

We identified a number of QTL associated with various agronomic and seed composition traits in soybean including seed yield, plant height, maturity, protein concentration, oil concentration, seed size, storage protein fractions, amino acid composition and nitrogen accumulation in seeds during reproductive stages. QTL associated with nitrogen accumulation, 7S and 11S storage protein fractions, and amino acid composition are being reported for the first time through this study. Some of the QTL associated with other traits like protein concentration, oil concentration, seed size, plant height, maturity, lodging and seed yield were verified from this study whereas others were new. There were two new QTL (near Satt042 and Satt570) detected associated with seed filling period in the present population. QTL near Satt292 was also a new associated with plant height whereas the one near Satt184 was verified in the present population. Among the two QTL (near Satt225 and Satt593) associated with lodging, the one near Satt593 was novel. Among six QTL linked to the molecular markers Satt042, Satt252, Satt263, Satt292, Satt293 and Satt591 associated with maturity in this population, the QTL near Satt263, Satt292 and Satt293 were new whereas the remaining three were verified. Among the three QTL (near Satt042, Satt076 and Satt412) associated with seed yield, only one near Satt412 was a new QTL.

Two QTL (near Satt274 and Satt570) were detected as protein QTL but the one near Satt570 was the only novel QTL. There were seven QTL

linked to the molecular markers Satt274, Satt293, Satt317, Satt420, Satt479, Satt562 and Satt570 associated with oil concentration in this population. QTL near Satt274, Satt420 and Satt479 were new ones that are identified to be associated with oil concentration. There were eight QTL near Satt002, Satt147, Satt184, Satt185, Satt255, Satt268, Satt459 and Satt540 associated with seed size in this population. Among those the one near Satt147, Satt255 and Satt459 were novel seed size QTL. The verified QTL are robust and can be used either in marker-assisted selection or positional cloning. The new QTL associated with protein quality and agronomic traits provide new information, which needs to be validated in other populations and environments.

The present research also provided new insight for future research. We were able to release a new germplasm for its total sulfur containing amino acids that is up to the standard of FAO/WHO. This proved the concept of possibility of developing a genotype capable to meet the FAO/WHO standard by conventional breeding. However, there is no information about the sulfur containing amino acid levels in the USDA soybean germplasm collection. A need exists for that germplasm bank to be analyzed and documented. This would be helpful in choosing the parents for crossing program aiming to improve the amino acid profile in soybean. Furthermore, unlike past findings, we had a positive correlation between 7S and 11S fractions of storage protein. This leads us to conclude that,

depending upon the genotype, the typical negative correlation between 7S and 11S may be overcome. This is possible if sulfur containing amino acids are present in sub-units of the 7S fractions. Therefore an opportunity exists to analyze the 7S and 11S fractions and their subunits for sulfur containing amino acid composition. This may lead to identification of specific genotypes that contain elevated level of sulfur containing amino acids. Validation of the QTL associated with new traits such as nitrogen accumulation, amino acid composition and 7S and 11S fractions identified in the present study is necessary, leading immediate nutritional improvement in soybean. An important aspect of the QTL detected or validated from this study is future research on gene expression patterns. Studies on expression of genes under various environmental conditions would not only lead us to better understand the functional genomics related to a particular trait, but also help to identify specific isolines for protein fractions and amino acid profile improvement in a breeding program. Therefore, four areas are identified for future research: germplasm screening for sulfur containing amino acids, validation of protein quality QTL, analysis of a population for amino acid composition in the 7S and 11S fractions and their sub-units, and functional genomics of protein and amino acid composition in soybean.

VITA

Dilip Raj Panthee was born in Gulmi, Nepal on June 15 1964. He was raised in Dhurkot, Gulmi and went to primary school and Himalaya High School at Gulmi. He graduated from Himalaya High School in 1981. From there, he went to Institute of Agriculture and Animal Sciences, Paklihawa and Rampur Campus, and received B.Sc. in Agriculture with an elective in Genetics and Plant Breeding in 1990. He joined Department of Irrigation in Kathmandu as Asst. Agronomist after B.Sc. Ag. and worked for a year studying farming system under irrigated and unirrigated conditions in Eastern Nepal. He joined Lumle Agriculture Research Center (LARC) as a Seed Technologist in 1991 and worked in seed quality control of vegetable and cereal seeds including field inspection and seed analysis. His research involved minimizing seed storage loss caused by storage pests, seed production, seed dormancy and seed viability of vegetable and cereal seeds. He earned his MPhil in Plant Breeding and Crop Breeding from the University of Birmingham and Reading, UK in 1997 and worked as a Plant Breeder at LARC for a short time. He joined Institute of Agriculture and Animal Sciences, Tribhuvan University as a Lecturer in 1998 and taught Plant Biotechnology, Genetics and Plant Breeding. His research was on exploring genetic diversity of garlic available in Nepal.

Dilip enrolled at the University of Tennessee in 2001 as a Graduate Research Assistant and received a PhD in Plant Sciences in 2005. Dilip's future plan is to concentrate in crop improvement using conventional and molecular techniques.