Genetic Analysis of Quantitative Trait Loci Associated with Seed Sucrose Content Using Molecular Markers in an Interspecific *Glycine* Cross

Mine Cicek

Thesis submitted to the Faculty of the Virginia Polytechnic Institute and State University in partial fulfillment of the requirements for the degree of

Master of Science

in

Crop and Soil Environmental Sciences

Glenn R. Buss, Co-chairman

M.A. Saghai Maroof, Co-chairman

Richard E. Veilleux

July 28, 1997
Blacksburg, Virginia

Keywords: Carbohydrates, restriction fragment length polymorphism, QTL

Genetic Analysis of Quantitative Trait Loci Associated with Seed Sucrose Content Using Molecular Markers in an Interspecific *Glycine* Cross

Mine Cicek

(ABSTRACT)

Sucrose content is one of the important seed quality traits in soybean, especially for oriental soyfood production. However, little genetic information is available on this quantitative trait yet.

A previous study was conducted on seed sucrose content of soybean using a population of F_2-derived lines from an interspecific cross between an adapted high-sucrose (8.3%) *G. max* breeding line (V71-370) and a low sucrose (1.6%) *G. soja* plant introduction (PI407162). Nineteen marker loci, mapping to seven linkage groups (A1, A2, E, F, L1, I, and M), were significantly associated with seed sucrose content after screening 178 polymorphic genetic markers, including RFLPs, SSRs, RAPDs and morphological markers. The replicated field experiments were planted in 1993 and 1995.

The objective of my study was to evaluate QTLs associated with seed sucrose content utilizing an additional 153 $F_{2:3}$ families from the same cross.

DNA samples from the additional families were analyzed with the nineteen genetic markers associated with sucrose in the previous study. Sucrose data were obtained from seeds harvested from a field experiment conducted in 1995. Single factor analysis of variance results for the sucrose data obtained from the 153 $F_{2:3}$

families were compared to the 1995 data for the 144 $F_{2:3}$ families of the previous study.

Of the nineteen genetic markers significantly associated with seed sucrose content in the previous study, seven were also significantly associated in this study. These genetic markers include sgA458a on linkage group A2, NBS61 on linkage group E, sgB164, R-B4a and sgB162 on linkage group L1, and R-B4b and sgA144 on linkage group I. The percent phenotypic variation explained by significant individual markers varied from 2.9 to 6.8% in the 153 $F_{2:3}$ families.

This study shows that seed sucrose content, a quantitative trait, may be improved using the molecular marker technology. Further research is necessary in different genetic backgrounds of *G. max* in order to implement these markers in a breeding program for selection.

Dedicated to my Parents

Sevgi and Vedat Sengur

ACKNOWLEDGEMENT

I express my appreciation to my advisors, Drs. G.R. Buss and M.A. Saghai Maroof, for their tremendous guidance, encouragement, and patience throughout my studies. Their scientific spirits promoted me to work harder and harder. I would also like to acknowledge the guidance and assistance received from Dr. R.E. Veilleux. I am honored to be able to have the chance to work with such outstanding scientists.

I also would like to express my sincere gratitude to those research scientists, students, Eastern Virginia Agricultural Research and Extension Center and Virginia Crop Improvement Association farm staff with whom I have enjoyed friendship and collaborations, namely: Dr. P. Chen, Dr. R. Biyashev, Dr. P.J. Maughan, Dr. Z. Liu, A. Hayes, Dr. G. Ma, P. Bagade, L. Barrack and Bruce Beahm. The financial support of the Montague Farms, Inc. is gratefuly acknowledged.

I express my sincere love for my caring parents, Sevgi and Vedat Sengur, whose support and encouragement have made my education possible. Lastly I give my greatest thanks to my loving and caring husband, Muzaffer, whose patience, understanding and scientific guidance during my graduate years have made my education feasible.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

INTRODUCTION

Soybean (*Glycine max* (L.) Merrill) is often called the miracle crop of the century, but is one of the oldest cultivated crops. Taxonomically, the soybean is classified in the Leguminosae family and genus *Glycine*. The genus *Glycine* subgenus *soja* consists of the cultivated soybean *Glycine max* (L.) Merrill and its wild progenitor *Glycine soja* (L.) Sieb. and Zucc. The soybean was domesticated in northeastern China and was introduced into the United States in the late 1700's. Most of the early introduced cultivars were only grown for forage. With the improvements in soybean breeding, it became an important grain crop in the U.S. in the 1940's. Today, it is third in production of grain after corn and wheat, and second in value after corn in the U.S. Soybean has been mainly improved by conventional breeding methods. Improvement most commonly involves crosses between cultivars to generate recombinant genotypes from which superior cultivars are selected (Fehr. 1987). After creating the F_1 hybrids, generations are advanced to near-homozygosity by self-fertilization. Inbred lines are then evaluated in replicated field trials at multiple locations. Lines possessing superior agronomically important traits are released as new cultivars.

Soybean cultivars grown in the United States have a narrow genetic origin. A few accessions have been used to derive the commonly grown soybean cultivars in the U.S. today. Pedigree analysis of cultivars of northern U.S. germplasm indicates that 88% of their genome was derived from just 10 accessions. The situation for southern germplasm is even more extreme, with 70% of the genome contributed by as few as seven accessions (Delannay et al., 1983). There is a lack of genetic diversity which may cause serious problems such as outbreaks of

disease epidemics in the future. Disease problems and drought have been reported as important causes for high yield loss in soybean production in the southern states (Sciumbato and Turnage. 1992). Most of the breeding programs have the objective to develop cultivars resistant to diseases, nematodes, and insects common to their production areas. The use of molecular markers could accelerate the genetic process in many soybean breeding programs (Skorupska et al., 1993).

***Glycine max* and *G. soja* both are diploid and carry the chromosome number 2n=40. They are cross fertile, but because of some of the undesirable characteristics of *G. soja* it is mostly used as a donor parent to transfer a gene of interest for a desirable trait to the recurrent parent. It may be backcrossed to the cultivated species to recover a desirable plant type.**

Soybean cultivars grown in the U.S. typically contain about 20% oil, 40% protein and 15% soluble carbohydrates (Hsu et al., 1973). It has been found that sucrose and the oligosaccharides, raffinose and stachyose, are the major carbohydrates in soybean seeds (Openshaw and Hadley. 1978). They make up more than 99% of the sugars in soybean seeds (Kawamura. 1967), and occur in a ratio of approximately 8:4:1 (w/w/w), respectively (Hymowitz et al., 1972b). It has been reported that sucrose content, which is much lower in seeds of *G. soja* than in seeds of *G. max* (Hymowitz and Collins. 1974), is positively correlated with oil content and negatively correlated with protein content (Hymowitz et al., 1972a).

Considering the economic importance of the soybean crop, genetic research on soybean has been increased tremendously in a short time. Soybean production started in the USA by using various cultivars introduced from other countries. These plant introductions were evaluated and the ones with high yield and good quality were released for commercial use. Crosses have been made between

desirable plant introductions for further selections. Today, with the help of traditional plant breeding procedures, dramatic improvements have been accomplished in soybean cultivars.

The interest of researchers in improved, more productive soybean cultivars have helped to increase soybean production (Smith and Huyser. 1987). The U.S. currently produces over 50% of the world soybean crop. Twenty five percent of total world soybean production enters the international trade market in the form of whole beans. There is also a sizable market for human soyfood products such as tofu, miso, natto, soy-milk, soysauce, etc. Ninety percent of the food quality soybeans are used in Japan for traditional soyfoods (Taira. 1990). Most U.S. soybean cultivars are poorly suited for soyfood production because they lack one or more of the seed characteristics considered important for soyfoods. Large-seeded soybeans (>20 mg/seed) with high protein and sucrose content and low oil content are preferred for tofu and miso production; on the other hand, small-seeded soybeans (<8 mg/seed) with high sucrose and low calcium content are preferred for natto production (Griffis and Wiedermann. 1990). A main objective of many soybean breeders is to improve the seed quality characteristics for soyfood as well as to maintain good agronomic traits.

These quantitatively inherited seed quality traits that are directly related with soyfood production are especially of interest for many soybean breeders. The main objective of these breeding programs is to produce new soybean cultivars that fit the quality criteria for these specific markets and also have acceptable agronomic traits.

Sucrose content is one of the important quality traits in soybean. The percent sucrose content of soybean seed is critical in soyfood production. There is not much previous work done on sucrose content. Sucrose is a quantitative trait and it

is impossible to evaluate the plants visually. Enzymatic determination of seed sucrose content is required after harvest. The easier, faster and more cost efficient way of screening plants for sucrose content is to use molecular marker technology.

The previous study on sucrose content of soybean seed in an interspecific *Glycine* cross was conducted by Maughan (1994). A segregating F_2 population was developed from interspecific hybridization between an adapted high sucrose *Glycine max* breeding line (V71-370) and a low sucrose *Glycine soja* plant introduction (PI407162) (Maughan et al., 1996). The experiment was replicated over two years with $F_{2:3}$ families in a randomized complete block design. Nineteen markers on seven linkage groups associated with seed sucrose content were identified (Maughan, unpublished data).

The objective of this study was to evaluate quantitative trait loci (QTLs) associated with seed sucrose content utilizing an additional 153 $F_{2:3}$ families from the same cross.

LITERATURE REVIEW

MOLECULAR MARKERS IN CROP IMPROVEMENT

Until the recent advances in molecular genetics, breeders have been improving agronomically important quantitatively inherited traits by conventional breeding methods based on phenotypic selection. In recent years, molecular biology has provided tools suitable for rapid and detailed genetic analysis. The most fundamental of these tools are DNA markers that detect differences in the genetic information carried by different individuals. The well-known application of DNA markers is in the construction of genetic maps, which are used to determine the location of genes on chromosomes that are related to agronomically important traits. By knowing the location of a gene, one can determine its presence by using nearby DNA markers, without actually observing the phenotype, which can save a lot of time and money, especially for quantitatively inherited traits.

There are different types of markers such as morphological markers, isozymes and DNA markers. A genetic marker associated with a trait of interest must be polymorphic between the parents and must be reproduced in the progeny in order to be useful. Morphological markers and isozymes have been used since the 1970's. In 1980, Botstein et al. reported that it could be possible to find a lot of genetic markers by studying the DNA molecule itself (Botstein et. al., 1980). DNA markers have advantages over morphological markers and isozymes. Greater numbers of DNA markers can be found whereas other markers are more limited. There are different types of DNA markers for different purposes. Below are the most widely used DNA markers for plants, animals, and humans:

Restriction Fragment Length Polymorphism (RFLP):

Restriction fragment length polymorphisms (RFLPs) are DNA markers which are fragments produced by restriction enzyme digestion that differ in length from individual to individual with different genotypes. The differing sizes of the DNA fragments may result from point mutations, deletions, insertions or transpositions. The fragments with different sizes can be analyzed first by seperating on agarose gel electrophoresis and then visualizing by autoradiography. RFLP markers are codominant and highly polymorphic.

Microsatellites or Simple Sequence Repeats (SSRs):

Microsatellites or simple sequence repeats (SSRs) have been described as an alternative source of DNA markers (Tautz. 1989). SSR is a polymerase chain reaction (PCR) based marker. The microsatellite loci consist of tandemly repeated sequences which are usually two to four nucleotides in length and are flanked by conserved DNA sequences. The primers complementary to these flanking regions can be used to amplify the locus by PCR. The PCR reaction includes one ^{32}P-labeled nucleotide to allow visualization of the amplified product by autoradiography after polyacrylamide gel electrophoresis. SSRs are highly polymorphic and codominant markers.

Random Amplified Polymorphic DNA (RAPD):

Random amplified polymorphic DNA technique is another PCR-based marker which is based on the amplification of random DNA segments with single primers of arbitrary nucleotide sequence. These polymorphisms are detected as DNA segments that amplify from one parent but not the other through examination of an ethidium bromide-stained agarose gel. Most of the RAPD markers are dominant

which is one of the disadvantages of this technique since it is not possible to distinguish heterozygotes from dominant homozygotes (Williams et al., 1990). This provides less information relative to markers which are codominant, such as RFLPs. On the other hand, RAPD markers have several advantages over other molecular marker techniques; no preliminary work is required such as digestion, blotting and hybridization, and only a small amount of DNA is required for analysis. RAPD markers are mostly used in screening genetic variability and for taxonomic purposes (Ragot and Hoisington. 1993; Lin et al., 1996).

Amplified Fragment Length Polymorphism (AFLP):

The amplified fragment length polymorphism (AFLP) technique is based on detection of genomic restriction fragments by polymerase chain reaction (PCR) amplification. The AFLP technique displays mostly presence or absence of restriction fragments, similar to RAPD markers, rather than length differences as in the RFLP technique, but rare codominant AFLP markers do occur. The number of fragments that can be analyzed in one reaction is typically 50-100 restriction fragments. This AFLP technique is powerful because the reliability of the RFLP technique is combined with the power of the polymerase chain reaction (PCR) technique (Vos et al., 1995).

MOLECULAR MARKER STUDIES ON SOYBEANS

Soybean breeders originally evaluated the existing and incoming plant introductions and then simply released the best ones to farmers as cultivars prior to the 1940's. Modern breeders search for potential cultivar releases by evaluating many recombinant lines that originate from the selfed progeny of hybridized parents. The genetic gain in yield is still the most important factor in

breeding programs. However, high yield potential must be complemented with genes that make plants less susceptible to stress which may cause yield loss. The basic challenge that the breeders face is to identify superior genotypes. Molecular genetics can contribute in this aspect by using genetic maps and marker-assisted selection. The soybean breeder can identify a recombinant's genotype by using molecular markers. The selection of parents is a very important issue in breeding programs. Molecular markers make it possible to see the diversity in breeding lines at the DNA level and to evaluate the potential parents.

The marker-assisted approaches are based on genetic linkage between a polymorphic marker locus and any locus controlling a trait of interest to the breeder. It is good to see how new technology can be integrated with conventional breeding programs. In the near future it may be a routine approach in cultivar improvement to use marker-assisted selection to save time, effort and money (Specht and Graef. 1996).

Genetic maps have been constructed for many plant species including maize (Helentjaris. 1987), rice (McCouch et al., 1988), barley (Huen et al., 1991; Kleinhofs et al., 1993), tomato and potato (Tanksley et al., 1992), sorghum (Ragab et al., 1994), cultivated oat (O'Donoughue et al., 1995), and alfaalfa (Kiss et al., 1997). Until recently, most of these maps were based on morphological and biochemical markers, but these maps have been extended widely by DNA-based markers.

Soybean genetic map development has proceeded slowly compared to genetic maps of other agronomically important crops. This was due to the lack of cytogenetic markers and proper genetic stocks. The application of molecular genetic mapping techniques has allowed development of a detailed soybean genetic map in a short time. The first efforts in constructing a soybean RFLP map

was reported in 1988. Apuya et al. (1988) showed that RFLP markers for *Glycine max* were useful in the construction of a genetic map of soybean. They identified RFLPs between 'Minsoy' and 'Noir', which are genetically widely distant cultivars of soybean *G. max* (L.) Merrill. Of 27 RFLP markers analyzed in 60 F_2 progeny plants. 11 markers were contained in four linkage groups.

In the early 1990's, studies were conducted to identify molecular markers associated with quantitative trait loci. Keim et al. (1990) constructed a molecular genetic linkage map and identified associations between molecular markers and quantitative trait loci. One hundred fifty RFLPs were used to identify genetic linkages in an F_2 segregating population from an interspecific cross between *Glycine max* and *G. soja*. Twenty six linkage groups containing 1200 recombination units (cM) were reported. Significant associations between RFLP markers and QTL were detected for eight of nine traits evaluated including leaf width, leaf length, stem diameter, stem length, etc. No significant associations were detected for internode length. The variation explained by each marker associated with a QTL was greater than 16% (R^2).

Restriction fragment length polymorphisms were mostly used to construct molecular genetic maps for soybean (Apuya et al., 1988; Keim et al., 1990). However, other DNA markers, random amplified polymorphic DNA (RAPD) and simple sequence repeat (SSR) DNA, were reported as genetic markers that can be placed in genetic maps in relation to other DNA markers and construct a more saturated soybean genetic map (Williams et al., 1990; Akkaya et al., 1992). Williams et al. (1990), successfully introduced a new DNA polymorphism assay (RAPD) which is very useful in mapping genomic regions which are not accessible to RFLP analysis due to highly repetitive DNA sequences. They confirmed the map positions of RAPD markers by RFLP analysis. Another type of

DNA marker, simple sequence repeat (SSR), was reported by Akkaya et al. (1992). They identified the presence and degree of simple sequence repeat DNA polymorphism in soybean. The highly polymorphic nature of SSRs in soybean is very important because RFLP polymorphism has been difficult to detect in some cases.

Several soybean genetic maps have been published already including hundreds of DNA-based markers (Lark et al., 1993; Rafalski and Tingey. 1993; Shoemaker and Olson. 1993; Akkaya et al., 1995; Mansur et al., 1996; Keim et al., 1997). Markers from these maps have been used for the identification of QTLs for agronomically important traits including disease resistance.

Shoemaker et al. (1992), using previously mapped clones, identified a set of DNA markers that had high probability of detecting polymorphisms among soybean genotypes. Using these selected markers they demonstrated that a detailed map can be used in pedigree analysis to determine the ancestral sources of markers in different cultivars.

Many different quantitative traits have been studied in soybean. Mansur et al. (1993a,b) developed a segregating population and a recombinant inbred population from a cross between two soybean cultivars: 'Minsoy' and 'Noir'. They analyzed 15 traits including reproductive, morphological and seed traits of soybean and found that 11 of the 15 traits were localized in 6 linkage groups. The morphological and developmental traits tended to be clustered in three intervals, two of which were also associated with seed yield. Seed oil content was separated from all of them. In this segregating population they presented the evidence of transgressive segregation. In the recombinant inbred population, lines showing an extreme phenotype were selected and the DNA samples were bulked for each trait analyzed. The results confirmed the linkage groups identified by their previous

interval mapping in a smaller segregating population. The advantage of recombinant inbreds over segregating population is the fact that the extremes can be identified and the linkages can be determined more rapidly and precisely (Mansur et al., 1996). Using the same recombinant inbred population, the epistatic expression of quantitative traits in soybean has been determined and inverse correlation between seed oil content and protein content has been reported (Lark et al., 1994).

Two segregating populations derived from crosses between 'Young' and PI416937 and PI97100 and 'Coker 237' were evaluated with RFLP markers to identify QTLs related to seed protein and oil content (Lee et al., 1996c). They found a negative correlation between seed protein content and oil content which agrees with earlier studies reported by Diers et al., (1992) and Lark et al., (1994). In a previous study on PI97100 x 'Coker 237', Lee et al. (1996 a, b), reported QTLs for plant height, lodging and maturity. They also constructed a genetic linkage map for this population covering about 1600 cM on 23 linkage groups. Brummer et al. (1997) reported the latest study on mapping QTL for seed protein and oil content in soybean populations. They identified environmentally stable QTLs that were sensitive to genetic backgrounds. All of eight populations studied had at least one stable QTL for protein after three years replication. No stable QTLs were identified in two of the populations for oil content. The markers associated with these stable QTLs could have potential in breeding programs to alter the populations for these traits.

Another group of researchers constructed an RFLP map from a soybean population derived from the cross of the same 'Young' and PI416937 lines to identify QTLs associated with water use efficiency which is an important trait that has been related to drought tolerance of soybean plants (Mian et al., 1996). They

identified a total of four RFLP markers associated with water use efficiency explaining 38% of the variability. Leaf ash trait was also studied in the same population which is found to be negatively correlated with water use efficiency. Six RFLP markers explaining 53% of the variability were reported for leaf ash trait in this population.

Two studies have been conducted on assessing the level of genetic variation in cultivated and wild soybeans using simple sequence repeats (SSRs) (Maughan et al., 1995) and amplified fragment length polymorphism (AFLP) (Maughan et al., 1996a). These authors reported a greater diversity for wild soybeans in both studies. They detected 5 to 21 alleles per locus with SSRs which are the largest numbers of alleles reported in soybean, highlighting the importance of SSR markers for low genetic diversity cultivars. The latter study showed the power of AFLP analysis making this class of markers very useful in genetic investigations.

A genetic map of the interspecific cross between a cultivated soybean *G. max* (V71-370) and plant introduction *G. soja* (PI407162) was constructed by Maughan et al., 1994). A total of 91 genetic markers (84 RFLP, 4 SSR and 3 RAPD) out of 229 screened was polymorphic between the parents of this cross. A genetic map of this interspecific cross was used in a study to identify quantitative trait loci controlling seed-weight (Maughan et al. 1996b). They identified 3 and 5 markers controlling seed-weight in the F_2 and $F_{2:3}$ generations, respectively, in this interspecific soybean F_2 population. These markers were significantly associated with seed-weight and their interactions explained 50% and 60% of the phenotypic variation in the F_2 and $F_{2:3}$ generations, respectively. In a subsequent study, the genetic linkage map constructed on this population was expanded. Three hundred and forty-nine genetic markers were screened for polymorphism and 178 of them were polymorphic between the parents of this population. This

improved map covered 1589 cM on 25 linkage groups with an average distance of 9.6 cM. This map was used in a study to identify QTLs for seed sucrose content in the same interspecific soybean F_2 population. A total of 19 molecular markers, on seven linkage groups (A1, A2, E, F, M, L1, and I) was found to be significantly associated with seed sucrose content from two years of study (Maughan et al., manuscript in preparation). These genomic regions explained 53% of the total variation for sucrose content in this population.

DNA markers have also been used to analyze pest resistance of soybean. Resistance to soybean cyst nematode (SCN) (Concibido et al., 1994. 1996a; Webb et al., 1995) and soybean mosaic virus (Yu et al., 1994. 1996) have been studied extensively. Concibido et al. (1994) using three segregating F_2 populations and different races of SCN reported a total of four genomic regions that were associated with cyst nematode resistance. Only one of these genomic regions was common to all mapping populations and three others were race specific. Yu et al. (1994) identified one SSR and two RFLP markers closely linked to a resistance gene in a soybean segregating population constructed from a cross between PI96983 and Lee68. These three markers were used in a marker-assisted screening for classification of 67 soybean accessions for the SMV resistance. Forty-five were resistant, one was necrotic and 21 were susceptible among a total of 67 accessions which were grouped into 20 classes. The classes were constructed according to their multilocus variants at SMV resistance genes (Yu et al., 1996). Hnetkovsky et al. (1996) identified and located QTLs associated with field resistance to a fungal disease of soybean sudden death syndrome (SDS) in adapted soybean genotypes by using molecular markers.

MATERIALS and METHODS

Genetic Materials

The original experimental population used in this study was developed by Maughan (1994). A segregating population consisting of approximately 400 F_2 plants was developed from interspecific hybridization between an adapted large-seeded (24 g/100 seeds), high sucrose (8.3%) *Glycine max* breeding line (V71-370) and a small-seeded (1.8 g/100 seeds), low sucrose (1.6%)*Glycine soja* plant introduction (PI407162) (Figure 1). One hundred and fifty of the F_2 plants were grown in the greenhouse in the winter of 1993 at Virginia Polytechnic Institute and State University, Blacksburg, Va. Thirty $F_{2:3}$ seeds from each of 148 F_2 plants were planted in a randomized complete block design with two replications at the Virginia Crop Improvement Association's seed farm near Mt. Holly, Va in summer 1993 (Maughan. 1994). $F_{2:3}$ lines were planted in plots 120 cm long with 230 cm between rows. All plants in each plot were harvested in bulk when 90% of the plants reached maturity and were dried in large grain bags. Each plot was threshed separately after they all were dried. The experiment was repeated in 1995 using remnant seeds from 144 of the same 148 $F_{2:3}$ families. It was planted in June 1995, at the same location with a similar experimental design (Maughan, unpublished). These two sets of 148 (planted in 1993) and 144 (planted in 1995) $F_{2:3}$ families will be referred to as the J population, hereafter.

For the present study I used another set of 200 F_2 seeds, from the same interspecific cross as described above. F_2 seeds were planted in the field in 1994 at Warsaw, Va and F_2 plants were harvested individually. There was poor emergence due to heavy rains after planting and consequent soil crusting. Missing

hills were replanted in order to maintain the population size. $F_{2:3}$ seeds were harvested from a total of 153 F_2 plants. I will refer to this set of 153 $F_{2:3}$ families as W population. Thirty seeds per plot from each of the 153 $F_{2:3}$ families of the W population were planted along with the 144 $F_{2:3}$ families of the J population in 1995. The J and W sub-populations were planted in separate blocks which were randomized within replications. The parents of the cross were repeated four times in each replication for both populations in a randomized complete block design.

Quantitative Trait Analysis

Sucrose content was analyzed using finely ground flour samples of 50 g $F_{2:4}$ seeds harvested from each of the 153 $F_{2:3}$ families in the field. Each plot was analyzed separately and the average of the two plots for each $F_{2:3}$ family was used in genetic analysis. Seed sucrose content was determined by using a sucrose/D-glucose enzymatic assay kit (Boehringer Mannheim kit #139041). Analysis was done in two steps including determination of moisture in the ground samples and enzymatic determination of sucrose.

Determination of Moisture:

Aluminum pans with tightly fitting lids were used for determining the moisture of each sample. Dry weight of each pan and lid was recorded using an analytical balance after placing them in an oven (130 +/-3˚C) for an hour and allowing them to cool down to room temperature in a dessiccator. Two grams of well-mixed ground soybean sample was added to each pan and the weight of the pan and lid was recorded. Pans and lids were placed in an oven at 130˚C for another hour and dry weights of samples were recorded after they cooled to room

temperature in a dessiccator. Calculation of moisture content and sample dry weight was according to the formulae given below:

Calculation of moisture content:

A = weight of pan and lid (grams)
B = weight of pan, lid, and sample (undried; grams)
C = weight of pan, lid, and sample (dried; grams)

$$\% \text{ moisture} = \frac{(B-C)*100}{B-A}$$

Enzymatic determination of sucrose:

One gram of ground soybean for each sample was used for the enzymatic sucrose assay. Total sugars were extracted by shaking samples in 100 ml of 80% ethanol for 18-24 h at 300 rpm at room temperature in 8-ounce wide-mouth Nalgene® screw cap bottles. After shaking samples with ethanol. 1 ml of extraction solution was filtered through a 0.45 μm filter (Acrodisc LC PVDF 25 mm; Gelman) and evaporated to dryness with compressed air using a Reactitherm heating/evaporation unit set at 98˚C. Sugar extracts were then dissolved in 1 ml distilled deionized water and assayed for sucrose content using a sucrose/D-glucose enzymatic assay kit and following manufacturer's guidelines. The D-glucose concentration was determined before and after enzymatic hydrolysis:

Determination of D-glucose before inversion:

1.) D-glucose + ATP $\xrightarrow{HK}$ G-6-P + ADP

2.) G-6-P + NADP+ $\xrightarrow{G6P\text{-}DH}$ Gluconate-6-phosphate + NADPH + H+

The enzyme hexokinase (HK) catalyzes the phosphorylation of D-glucose by adenosine-5'-triphosphate (ATP) with the simultaneous formation of adenosine-5'-diphosphate (ADP) (1). The glucose-6-phosphate (G6P) is oxidized in the presence of glucose-6-phosphate dehydrogenase (G6P-DH) by nicotinamide-adenine dinucleotide phosphate (NADP) to gluconate-6-phosphate with the formation of reduced nicotinamide-adenine dinucleotide phosphate (NADPH) (2). The NADPH formed in this reaction is stoichiometric with the amount of D-glucose and is measured by means of its absorbance at 340 nm by a spectrophotometer.

Determination of D-glucose after inversion:

3.) Sucrose + H_2O $\xrightarrow{\beta\text{-fructosidase}}$ D-glucose + D-fructose

Enzymatic inversion takes place in the presence of ß-fructosidase (invertase) with the hydrolysis of sucrose to D-glucose and D-fructose (3). The determination of D-glucose after inversion is carried out simultaneously according to the principles explained above (1) (2). The sucrose content is calculated from the difference of the D-glucose concentrations before and after enzymatic inversion according to the formulae given below:

$C = \{(A_2-A_1)\text{sample} - (A_2-A_1)\text{blank}\} * 1.641$

C = Sucrose (mg/ml)

A_2 = Absorbance reading after enzymatic inversion

A_1 = Absorbance reading before enzymatic inversion

$S = C * 100 * 100 / (W_s * 1000)$

S = % sucrose

W_s = initial dry weight of sample (mg)

$$R = \frac{(mg/ml)_1 + (mg/ml)_2}{(0.50) * 2}$$

R = mean recovery of sucrose from standards

Scorr = S/R

Scorr = correct sucrose content of samples for recovery of sucrose

Two other traits of this cross, seed size and maturity, were analyzed. The maturity trait was scored as the date after Aug. 31 when 50% of the pods were mature color. Seed size trait was scored as the weight (g) of 100 seeds.

Molecular Marker Assay

DNA Isolation:

Young leaves from eight-week-old F_3 plants were cut as tissue samples for DNA extraction from the field. An equal amount of tissue was taken from each single plant and bulked by rows. Only one replication was sampled for DNA extraction. The tissues were transported the same day to the laboratory on dry ice and stored at -80˚C.

DNA from parental and $F_{2:3}$ lines was isolated according to the protocol described previously by Saghai Maroof et al. (1984). Freeze-dried leaf tissue

(0.75 g, dry weight) was ground with a mechanical mill and dispersed in 15 ml of extraction buffer (50 mM Tris, 0.7 M NaCl. 10 mM EDTA. 1% hexadecyl-trimethylammonium bromide, 0.1% 2-mercaptoethanol), and incubated at 65˚C for 60 min in a shaker bath. Ten ml chloroform/octanol, 24:1 was added and the solution was mixed by inversion and was centrifuged at 3200 rpm at 4˚C for 15 min. The aqueous phase was transferred and 2/3 volume of isopropanol was added and mixed by inversion. The precipitated pellet of DNA was hooked with a glass rod and transferred into a glass tube containing 20 ml 76% EtOH / 10 mM NH_4Ac overnight. The DNA pellet was dissolved in 10 mM NH_4OAc / 0.25 mM EDTA. The fluorometer reading was taken for each $F_{2:3}$ line DNA in order to measure the DNA concentration. Uncut DNA was run on 0.8% agarose gel to check its quality.

Southern Blot Analysis :

Eight μg of DNA were digested with each of four restriction enzymes (*EcoR* I, *EcoR* V, *Hind* III and *Dra* I) for 16-18 h at 37˚C and electrophoresed on 0.8% agarose gels with Tris-Acetate-EDTA buffer at 80-100 amps for 12-18 h.

Following electrophoresis, the DNA restriction fragments were depurinated by soaking in 0.2 M HCl for 7 min. The DNA was denatured in 0.2 M NaOH-0.6 M NaCl for 30 min and then neutralized in 0.5 M Tris-HCl (pH 7.5)-1.5 M NaCl for 30 min before transfer. The gel was laid on top of 3MM filter paper which was saturated with transfer buffer (25 mM $NaPO_4$, pH6.5). Immediately, the top surface of the gel was covered with a nylon membrane and overlaid with dry filter paper and a stack of paper towels. The transfer of DNA was performed overnight for about 18-20 h. The DNA fragments were permanently fixed to the nylon

membrane by baking in an oven at 80◦C (Southern. 1975). The membrane was prehybridized at 65˚C overnight.

Southern blots were hybridized overnight with approximately 25 ng of DNA insert, which were available in the laboratory, labeled with randomly primed ^{32}P-dCTP at 65˚C. Following hybridization, blots were washed four times, twice for 5 min at room temperature with 1xSSC and 0.1xSDS and twice for 10 min at 65˚C with 0.1xSSC and 0.1xSDS. After washing, the blots were exposed to Kodak X-ray film with intensifying screens at -80˚C for 5-7 days (Yu et al., 1994) .

Simple Sequence Repeat (SSR) Analysis:

In addition to several RFLPs, one SSR marker was reported to be associated with the QTL for sucrose content (Maughan. 1994). SSR procedures were as described by Yu et al. (1994). Briefly, a 20 μl PCR reaction contained 50 ng of genomic DNA, 0.1 mM of each primer. 10x reaction buffer, 3 mM $MgCl_2$, 200 mM each dATP, dGTP, and dTTP, 5 mM dCTP. 1.0 U of AmpliTaq DNA polymerase (Perkin-Elmer Cetus, Norwalk, Ct.), and 1 μCi [α-^{32}P]dCTP. Reaction mix was covered with mineral oil to prevent evaporation, and the reaction was processed at 94˚C for 3 min, at 55˚C for 2 min, and at 72˚C for 1.5 min for one cycle, followed by 30 cycles at 94˚C for 1 min, at 55˚C for 2 min, and 72◦C for 1.5 min, with a final extension step of 72˚C for 5 min. PCR products were denatured at 94˚C for 5 min. Five μl of each sample was loaded on a formamide polyacrylamide denaturing gel and separated at 1500-V constant power in 1x TBE (Tris-Borate-EDTA) running buffer, using a DNA sequencing unit (Model STS-45, IBI, New-Haven, Ct.). Gels were immediately covered with plastic wrap and exposed to X-ray film for 1 h.

Randomly Amplified Polymorphic DNA (RAPD) Analysis:

The RAPD assay is based on the use of short, random-sequence oligonucleotides as primers for the amplification of randomly-selected segments of the target genome. PCR amplification was performed in 25 μl reaction containing 25 ng of genomic DNA, 0.2 mM primer (Operon kits B and O, Alameda, CA), 200 mM each of dATP, dGTP, dTTP, and dCTP, 2.5 mM $MgCl_2$. 10x reaction buffer. 1U Taq DNA polymerase (Promega, Madison, WI). Reaction mix was covered with mineral oil to prevent evaporation, and the reaction was processed at 95˚C for 3 min for one cycle, followed by 45 cycles at 94˚C for 1 min, at 35˚C for 1 min, and 72˚C for 2 min. Amplification fragments were separated according to size on 1.4% agarose gels, stained with ethidium bromide and visualized under ultraviolet light (Williams et al., 1990).

Data Analysis:

The J population molecular marker and quantitative trait data were taken from the previous study conducted by Maughan (1994). Two sets of 1993 and 1995 J population data and one set of 1995 W population data were analyzed separately. The average values of two replications were used to analyze the sucrose data for both of the populations. Sucrose means, ranges, and normality for each population were determined by JMP IN program (Statistical Analysis Systems, Cary, NC). The means and standard errors for both parents were also calculated as well as the midparent values. The significance of each marker's association with sucrose was determined by one-way analysis of variance (ANOVA) using the computer program SAS (Statistical Analysis Systems, Cary, NC). Marker genotypes were classes and the analysis of variance was determined separately for each marker. Genotypic class means, R-square values and P-values were

calculated for each marker. Analyses were also conducted using the combined 1995 data of both sub-populations. Observed frequencies at marker loci were used to calculate χ^2 values in both sub-populations to test goodness of fit to the expected Mendelian ratios.

RESULTS and DISCUSSION

As indicated on page 14, molecular marker and sucrose data on the J population were collected by Maughan (manuscript in preparation; Maughan. 1994). These data sets are presented here only for comparison purposes with the results from the W population.

Phenotypic analysis:

The population size, sucrose content mean, range, sucrose content of the parental lines, midparental values and standard error of the parental means of the V71-370 and PI407162 interspecific cross are presented in Table 1.

Both populations showed a normal distribution as one would expect for a quantitative trait. A Shapiro-Wilks test for normality indicated that both populations were normally distributed (P>0.5) with a very slight negatively skewed distribution (Figures 2, 3, 4). The average sucrose content in all three data sets was significantly below the midparent value. These results suggest that there is partial dominance of the alleles for low sucrose content. Transgressive segregation was not observed in either population. In fact, the range of the F_2 progeny was considerably less than the range of the parents.

The mean sucrose content of the parental lines and the mean of the J population were somewhat lower in 1995 than in 1993 but the difference was not significant (Table 1). Data for two years were available only for the J population which indicated no significant genotype-by-year interaction (Maughan, unpublished data). The mean of the W and J populations also were in good agreement for sucrose content when both were grown in 1995. The 95% confidence limits of the

means for the two populations were overlapping indicating that these two subsets are not significantly different at P=0.05 (Figures 3, 4).

Two other traits of this cross, seed size and maturity, were analyzed in order to make sure that these two subsets were truly representative random samples of the same population (Table 2). For maturity, parental values as well as progeny means were similar for the two populations. Seed size means from the two populations as well as the parents were also quite similar (Table 2). These observed similarities indicate that the two subsets are representing the same population.

Genetic analysis:

Previous interval-mapping analysis of seed sucrose content QTLs in the J population were presented under the section: Molecular Marker Studies on Soybeans (page 12). Locations and genotypic class means for the 19 markers significantly associated with variation for seed sucrose content in the J population in 1995 (Maughan, unpublished) are presented in detail in Table 3.

There was no significant genotype-by-year interaction between 1993 and 1995 data (personal communication), so discussion of data from the J population will be limited to the 1995 data, since that is more directly comparable to the W population data. Percent phenotypic variation explained by individual markers, which are significantly associated with seed sucrose content at each QTL, varied from a low of 5.1% to a high of 12.3% for sucrose content in J population in 1995 (Table 3).

Based on the J population data (Maughan, unpublished), the 19 markers significantly associated with sucrose content were tested on the W population. The results are shown in Table 4. It is clear that none of the markers were found

to be as significant as they were in the J population. In fact, only seven of the 19 markers (sgA144, sgA458a, R-B4b, R-B4a, sgB164, NBS61 and sgB162), showed significant association with sucrose content at P<0.1. The percent phenotypic variation explained by significant individual markers varied from 2.9 to 6.8% in seed sucrose content in the W population.

Markers sgA458a and sgA144 on linkage groups A2 and I, respectively, seemed to be the strongest, the most reliable ones. They explain 6.1 and 6.8% of the total phenotypic variation, respectively. The other apparently important linkage group, L1, contains three significant markers, sgB162, R-B4a and sgB164, associated with seed sucrose content. Markers sgB164 and R-B4a were co-segregating and sgB162 was 24 cM away from them. The most significant marker locus in linkage group L1 was sgB164 based on its R-square value of 4.4%. However, R-B4a could also be as useful given that it was co-segregating with sgB164. The other marker, sgB162, being 24 cM away could be detecting another QTL or could be an indication of a single QTL somewhere near the morphological marker L1, which need to be clarified with more future study. Another significant marker NBS61 with a similar R-square value of 4.5% is covering a genomic region on linkage group E. None of the markers on linkage groups A1, F, and M were significant. This nonsignificance could be an indication to the fact that there are more markers on these chromosomal regions that we did not detect just because we only tested 19 markers. An RFLP marker sgB170 could be a good example for this. It was highly significant in the J population in 1995 but not in 1993 (Maughan et al., unpublished). Also, it was not significant in the W population. Nevertheless, it appears likely that sgB170 is associated with seed sucrose content. The homozygous *G. soja* genotypic mean for the sgB170 locus is higher than the homozygous *G.max* genotypic mean in all

populations which is the opposite of all the other markers, including sgT169 which is also in linkage group A1 (Table 4). Also, it is highly significant when the J and W populations are analyzed together. This, together with the fact that sgT169 and sgB170 are 99.7 cM apart seems to indicate that linkage group A1 contains two sucrose QTLs (Figure 5).

The same analysis of variance was also conducted on the data from the combined populations in 1995 (Table 5). The percent variation explained by each individual marker, except marker A458a, had intermediate values between the two sub-populations as might be expected. The exceptional marker, A458a, on linkage group A2 had an R-square value almost equal to the J population and higher than the W population, which suggests that this marker is considerably important. It's R-square value was second only to that of sgA144 in the analysis of the W population and in the combined analysis.

Significant agreement between the J and W populations when analyzed individually is only observed in seven markers (sgA458a, NBS61, R-B4b, sgA144, sgB164, R-B4a, and sgB162) on four linkage groups, including A2, E, I, and L1. These markers appear to be highly significant and strongly reliable for marker assisted selection programs. The other three linkage groups, A1, F, and M, have a considerably important disagreement. No studies have been published related to agreement or lack of agreement among sub-populations; however, the disagreement among the subsets of lines in 1995 seemed to be more than random variation. Different factors could be the cause of this disagreement between the sub-sets of lines.

Marker loci that significantly deviate from expected genotypic ratios ($P<0.05$) could explain the disagreement of some of the QTLs between the two sub-sets.

According to the χ^2 test results; four marker loci in the "J' population (sgA487, sgA144, sgK227, NBS61) were reported by Maughan to deviate significantly from their expected ratios and all four were significantly associated with seed sucrose content (Table 5). In the W population, three marker loci (sgA487, sgB170, R-B4a) were observed to deviate significantly from their expected ratios, and only one of them, sgA487, was also deviant in the J population (Table 6). Interestingly, that marker shows a good fit to a 1:2:1 ratio when the two populations are combined. This shows that a larger size population could give statistically a better fit to the Mendelian segregation ratio.

New strategies need to be implemented in order to improve markers for the sucrose trait. Three ideas can be evaluated. One is to repeat the experiment to get a second year's data for the W population. A second idea is to add new markers in genomic regions containing putative QTLs for seed sucrose content for a more saturated soybean map. This would increase the probablity of affecting seed sucrose content. Third is to implement a similar experiment on a recombinant inbred population of the same interspecific cross. This would give us another population to compare with the original populations. More importantly, it gives a stable population which can be grown in different environments without any change in the genotypes from generation to generation.

It is clear from the data presented that the two subsets of lines are good representative random samples of the same population. Several potential QTLs associated with seed sucrose content are identified. The low level of variation explained by the majority of individual QTLs confirms the quantitative nature of inheritance for this trait. Such QTLs could be useful tools for marker-assisted breeding programs. However, it has been reported that major QTLs need to be

studied in different genetic backgrounds before marker-assisted breeding programs are implemented because QTLs can have significant changes in their importance among populations and environments (Lee et al., 1996). This is particularly important in this interspecific populations, since *G. soja* exhibits more polymorphism than *G. max* (Maughan et al., 1995). It is possible that the sucrose QTLs found in this study will not be polymorphic in crosses among *G. max* lines. Also, the reason for the disagreement on the significance of some of the markers between the subsets is yet to be clarified. Improvement of this project by some of the new strategies mentioned above would provide useful information for soybean breeders trying to manipulate the quantitative trait, seed sucrose content.

Table 1. Population size, sucrose content mean, range of the F_2 progeny, sucrose content values for parental lines, midparent values and standard errors of the parental means for subsets J in 1993 and 1995 and W for 1995.

Population	N	Mean	Min	Max	V71-370	SE [a]	Midparent	PI407162	SE [a]
J (1993)[b]	148	4.2	2.4	5.8	8.5	0.39	5.1	1.6	0.16
J (1995)[b]	144	3.6	2.1	5.1	7.8	0.41	4.7	1.5	0.21
W (1995)	153	3.8	2.3	5.4	8.3	0.32	5.0	1.7	0.15

[a] Standard error of the parental means
[b] Sucrose data taken from Maughan et al. (unpublished data)

Table 2. Summary of the comparison between J and W populations for sucrose, seed size and maturity traits in 1995.

	Sucrose (%)		Seed size (g/100 seeds)		Maturity[a]	
	J[c]	W	J[d]	W	J	W
N	144	153	149	164	148	167
$F_{2:3}$ Mean	3.6	3.8	5.3	5.6	44	42
Min	2.1	2.3	3.2	3.1	28	29
Max	5.1	5.4	8.4	8.3	59	60
V71-370	7.8	8.3	24.2	24.6	46	46
Midparent	4.7	5.0	12.9	13.0	41	40
PI407162	1.5	1.7	1.5	1.5	35	34
SEM[b] (V71-370)	0.1	0.1	0.4	0.6	0.3	0.3
SEM (PI407162)	0.1	0.1	0.1	0.5	0.7	0.8

[a] Date after Aug. 31 when 50% of the plants are mature color
[b] SEM = Standard error of the mean
[c] Sucrose data taken from Maughan (unpublished data)
[d] Seed size data taken from Maughan et al. (1996)

Table 3. Locations and genotypic class means (% sucrose) for markers associated with variation for seed sucrose content in J-population (1995)[g].

LG[a]	Marker	MM[b]	MS	SS	R-square	P>F[c]
A1	sgA487	3.9	3.5	3.5	7.6	0.005
	sgT169	3.8	3.6	3.5	5.8	0.018
	sgB170	3.4	3.7	3.8	7.3	0.006
A2	sgA458a[f]	4.0	3.6		7.9	0.001
	sgA136	3.8	3.7	3.4	9.4	0.001
	sgA486	3.7	3.8	3.4	9.5	0.002
	sgT153	3.8	3.7	3.3	8.7	0.002
E	sgA963	3.8	3.6	3.5	6.2	0.016
	NBS61	3.9	3.7	3.4	5.1	0.032
F	R-O20[ef]	3.9	3.6		6.2	0.006
	sgA186	3.9	3.7	3.4	10.0	0.001
M	Sc514[d]	3.9	3.5	3.6	7.5	0.008
L1	sgB164	3.9	3.7	3.4	9.8	0.001
	R-B4a[ef]	3.9		3.6	7.5	0.011
	sgA23	3.9	3.7	3.4	9.8	0.001
	sgB162	3.7	3.7	3.3	7.5	0.005
I	R-B4b[ef]	3.8	3.4		9.0	0.005
	sgA144	3.9	3.7	3.2	12.3	0.001
	sgK227	3.9	3.6	3.2	10.9	0.001

[a] Linkage group designations according to the USDA-ISU soybean map
[b] MM - homozygous *G. max*; MS - heterozygous; SS - homozygous *G. soja*
[c] P-values from one-way ANOVA
[d] SSR marker
[e] RAPD markers
[f] Dominant marker loci
[g] Sucrose and molecular marker data taken from Maughan et al., unpublished.

Table 4. Locations and genotypic class means (% sucrose) for markers associated with variation for seed sucrose content in W-population (1995).

LG[a]	Marker	MM[b]	MS	SS	R-square	P>F[c]
A1	sgA487	4.0	3.8	3.6	3.2	0.131
	sgT169	3.9	3.8	3.7	1.7	0.305
	sgB170	3.7	3.9	3.8	1.6	0.345
A2	sgA458a[f]	4.0	3.7		6.1	0.007
	sgA136	3.9	3.7	3.7	1.7	0.275
	sgA486	3.9	3.7	3.7	1.5	0.345
	sgT153	3.9	3.7	3.6	2.4	0.230
E	sgA963	3.7	3.8	3.6	1.3	0.514
	NBS61	3.8	3.8	3.5	4.5	0.056
F	R-O20[ef]	3.7	3.8		0.4	0.444
	sgA186[f]	3.8	3.7		0.6	0.361
M	Sc514[d]	3.9	3.8	3.6	2.9	0.131
L1	sgB164	3.8	3.9	3.6	4.4	0.043
	R-B4a[ef]	3.9		3.7	2.9	0.041
	sgA23	3.8	3.8	3.7	1.2	0.433
	sgB162	4.0	3.7	3.7	3.4	0.096
I	R-B4b[ef]	4.0	3.7		5.5	0.005
	sgA144	4.0	3.8	3.5	6.8	0.006
	sgK227	3.9	3.7	3.8	1.7	0.318

[a] Linkage group designations according to the USDA-ISU soybean map
[b] MM - homozygous *G. max*; MS - heterozygous; SS - homozygous *G. soja*
[c] P-values from one-way ANOVA
[d] SSR marker
[e] RAPD markers
[f] Dominant marker loci

Table 5. Locations and genotypic class means (% sucrose) for markers associated with variation for seed sucrose content in combined populations J and W (1995).

LG[a]	Marker	MM[b]	MS	SS	R-square	P>F[c]
A1	sgA487	3.9	3.7	3.6	4.3	0.003
	sgT169	3.9	3.6	3.6	2.7	0.023
	sgB170	3.5	3.8	3.8	3.5	0.008
A2	sgA458a*	4.0	3.6		7.8	0.001
	sgA136	3.9	3.7	3.5	4.3	0.002
	sgA486	3.8	3.8	3.5	4.2	0.003
	sgT153	3.8	3.7	3.5	5.0	0.001
E	sgA963	3.9	3.7	3.6	4.3	0.003
	NBS61	3.8	3.8	3.5	4.9	0.002
F	R-O20*	3.8	3.7		0.7	0.186
	sgA186*	3.9	3.7		2.4	0.008
M	Sc514	3.9	3.7	3.6	3.3	0.011
L1	sgB164	3.9	3.8	3.5	6.1	0.001
	R-B4a*	3.8		3.6	3.9	0.003
	sgA23	3.9	3.7	3.5	4.1	0.003
	sgB162	3.8	3.7	3.5	4.2	0.003
I	R-B4b*	4.0	3.7		6.2	0.001
	sgA144	3.9	3.7	3.9	8.5	0.001
	sgK227	3.9	3.7	3.6	2.7	0.025

[a] Linkage group designations according to the USDA-ISU soybean map
[b] MM - homozygous *G. max*; MS - heterozygous; SS - homozygous *G. soja*
[c] P-values from one-way ANOVA
***** Dominant marker loci

Table 6. Genotypic segregation ratios at marker loci associated with sucrose in V71-370 x PI407162 in 1995.

		J POPULATION					W POPULATION				
Lg[a]	**Marker**	**MM**[b]	**MS**	**SS**	χ^2	**P>F**	**MM**	**MS**	**SS**	χ^2	**P>F**
A1	sgA487	49	57	36	7.9	<0.05	21	79	29	7.4	<0.05
	sgT169	47	62	32	5.2	0.1-0.05	33	81	31	2.1	0.5-0.1
	sgB170	39	67	35	0.6	0.9-0.5	35	55	47	7.4	<0.05
A2	sgA458a	25	111		3.2	0.1-0.05	35	84		1.2	0.5-0.1
	sgA136	34	66	42	1.6	0.5-0.1	40	74	36	0.2	0.9
	sgA486	23	72	41	5.2	0.1-0.05	38	75	32	0.7	0.9
	sgT153	28	79	33	2.7	0.5-0.1	31	59	33	0.3	0.9
E	sgA963	31	64	41	1.9	0.5-0.1	30	55	22	1.3	0.5-0.1
	NBS61	50	55	33	9.9	<0.01	34	74	20	6.2	<0.05
F	R-O20	24	99		2.0	0.5-0.1	34	107		0.1	0.9
	sgA186	31	72	40	1.1	0.9-0.5	46	107		2.1	0.5-0.1
M	Sc514	30	62	38	1.3	0.9-0.5	31	74	34	0.7	0.9
L1	sgB164	36	63	41	2.1	0.5-0.1	35	68	39	0.5	0.9
	R-B4a	61		26	1.1	0.5-0.1	80		63	27.7	<0.01
	sgA23	36	64	41	1.6	0.5-0.1	33	71	40	0.7	0.9
	sgB162	45	64	32	3.6	0.5-0.1	39	68	25	2.9	0.5-0.1
I	R-B4b	26	61		1.1	0.5-0.1	44	99		2.5	0.5-0.1
	sgA144	32	78	20	7.4	<0.05	38	83	26	4.4	0.1-0.05
	sgK227	39	82	20	8.9	<0.05	27	64	44	4.6	0.1-0.05

[a] Linkage group designations according to the USDA-ISU soybean map
[b] MM- homozygous *G. max*; MS- heterozygous; SS- homozygous *G. soja*

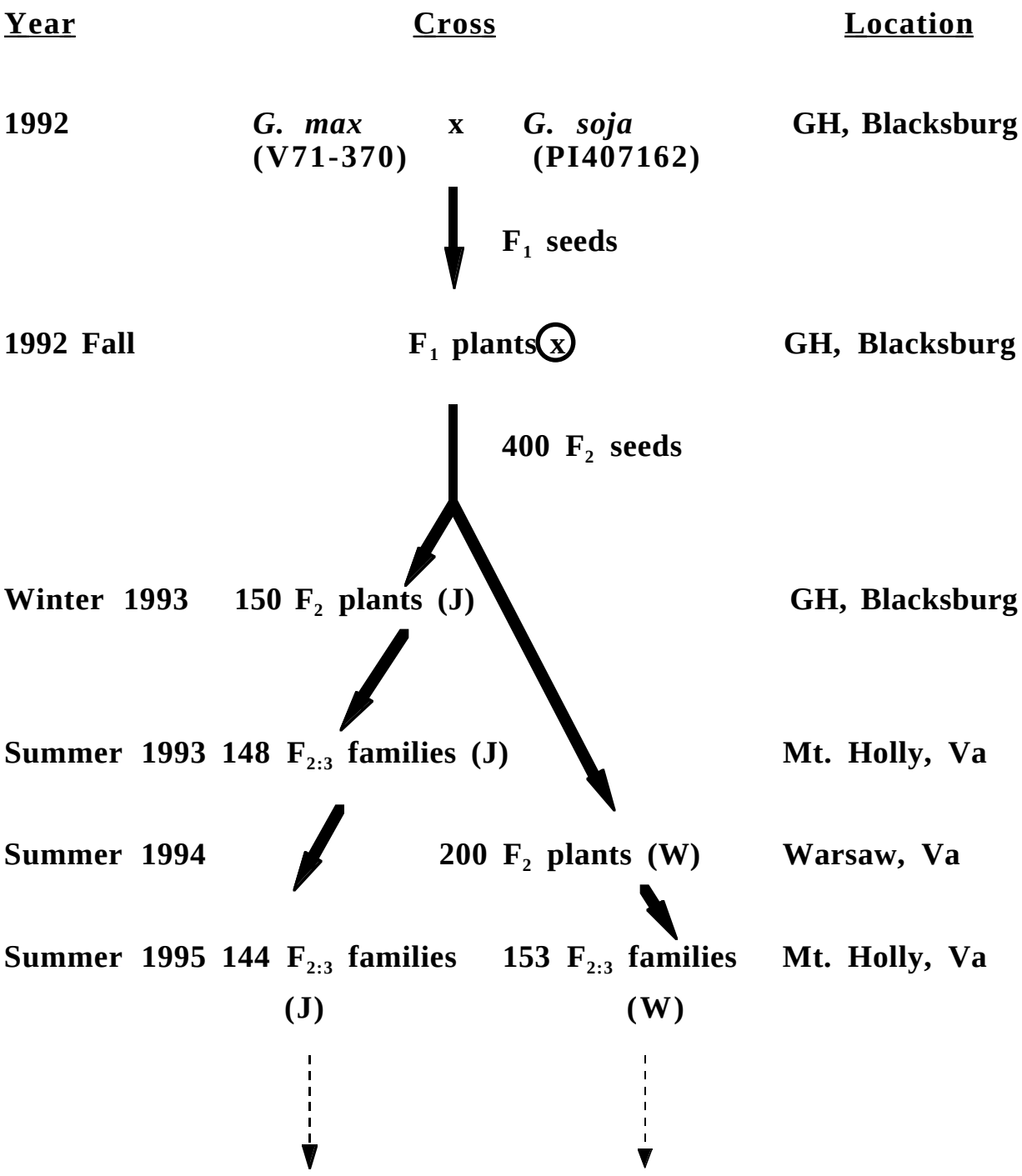
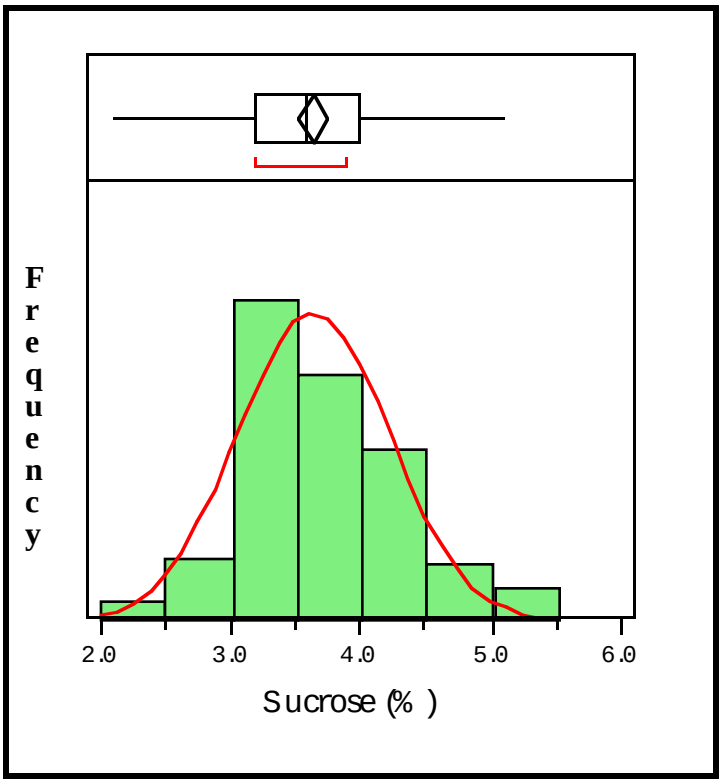


Figure 1. Development of the segregating F2 population from an interspecific cross between *G. max* and *G. soja*.



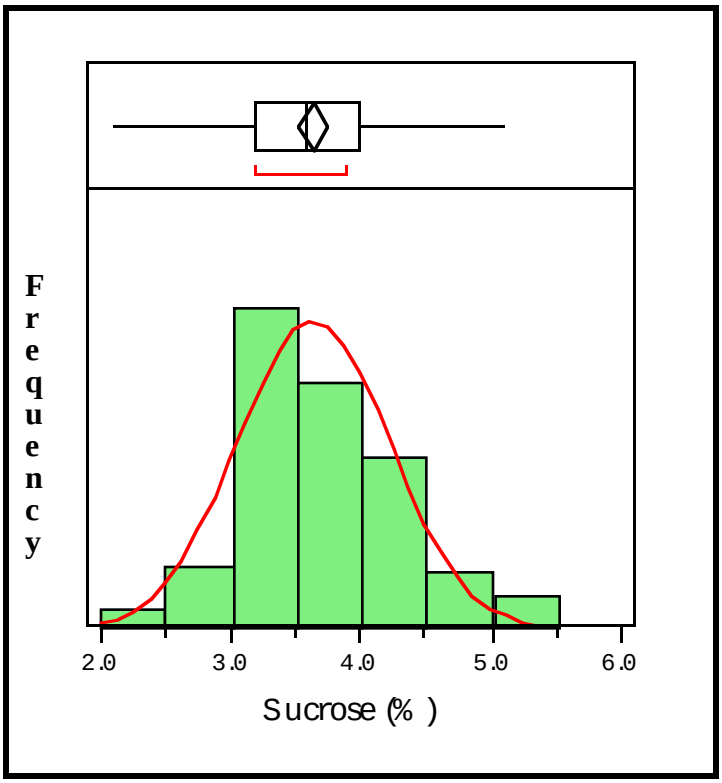
Quantiles

maximum	100.0%	5.8
	99.5%	5.8
	97.5%	5.5
	90.0%	4.9
quartile	75.0%	4.6
median	50.0%	4.1
quartile	25.0%	3.7
	10.0%	3.4
	2.5%	3.2
	0.5%	2.4
minimum	0.0%	2.4

Moments

Mean	4.2
Std Dev	0.6
Std Error Mean	0.5
Upper 95% Mean	4.3
Lower 95% Mean	4.1
N	148.0
Sum	614.6
Variance	0.4
Skewness	0.2
CV	14.6

Figure 2. Frequency distribution for percent seed sucrose content of the J-population grown at Mt. Holly in 1993.



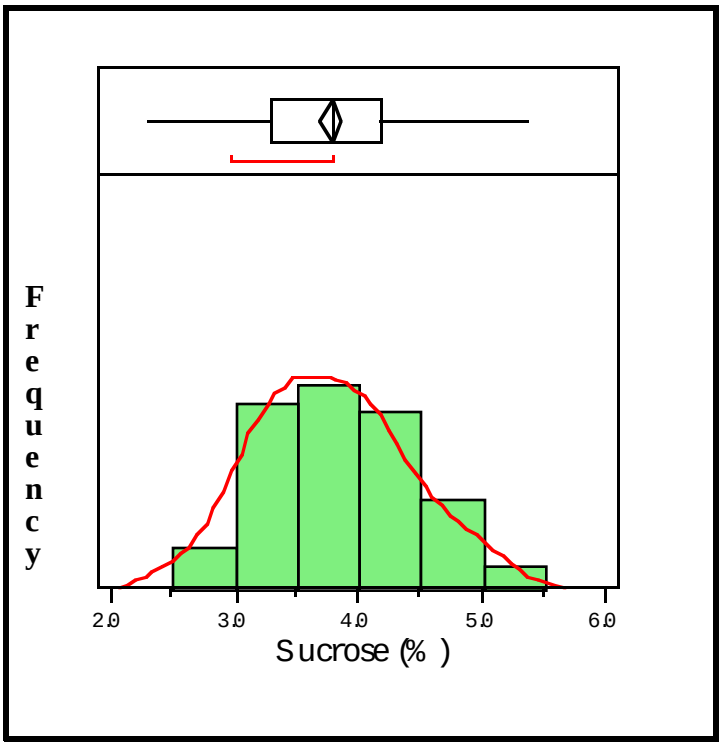
Quantiles

maximum	100.0%	5.1
	99.5%	5.1
	97.5%	5.0
	90.0%	4.5
quartile	75.0%	4.0
median	50.0%	3.6
quartile	25.0%	3.2
	10.0%	3.0
	2.5%	2.5
	0.5%	2.1
minimum	0.0%	2.1

Moments

Mean	3.6
Std Dev	0.6
Std Error Mean	0.1
Upper 95% Mean	3.7
Lower 95% Mean	3.5
N	144.0
Sum	524.2
Variance	0.3
Skewness	0.3
CV	16.1

Figure 3. Frequency distribution for percent seed sucrose content of the J-population grown at Mt. Holly in 1995.



Quantiles

maximum	100.0%	5.4
	99.5%	5.4
	97.5%	5.1
	90.0%	4.6
quartile	75.0%	4.2
median	50.0%	3.8
quartile	25.0%	3.3
	10.0%	3.0
	2.5%	2.6
	0.5%	2.3
minimum	0.0%	2.3

Moments

Mean	3.8
Std Dev	0.6
Std Error Mean	0.1
Upper 95% Mean	3.9
Lower 95% Mean	3.7
N	153.0
Sum	579.1
Variance	0.4
Skewness	0.2
CV	16.4

Figure 4. Frequency distribution for percent seed sucrose content of the W-population grown at Mt. Holly in 1995.

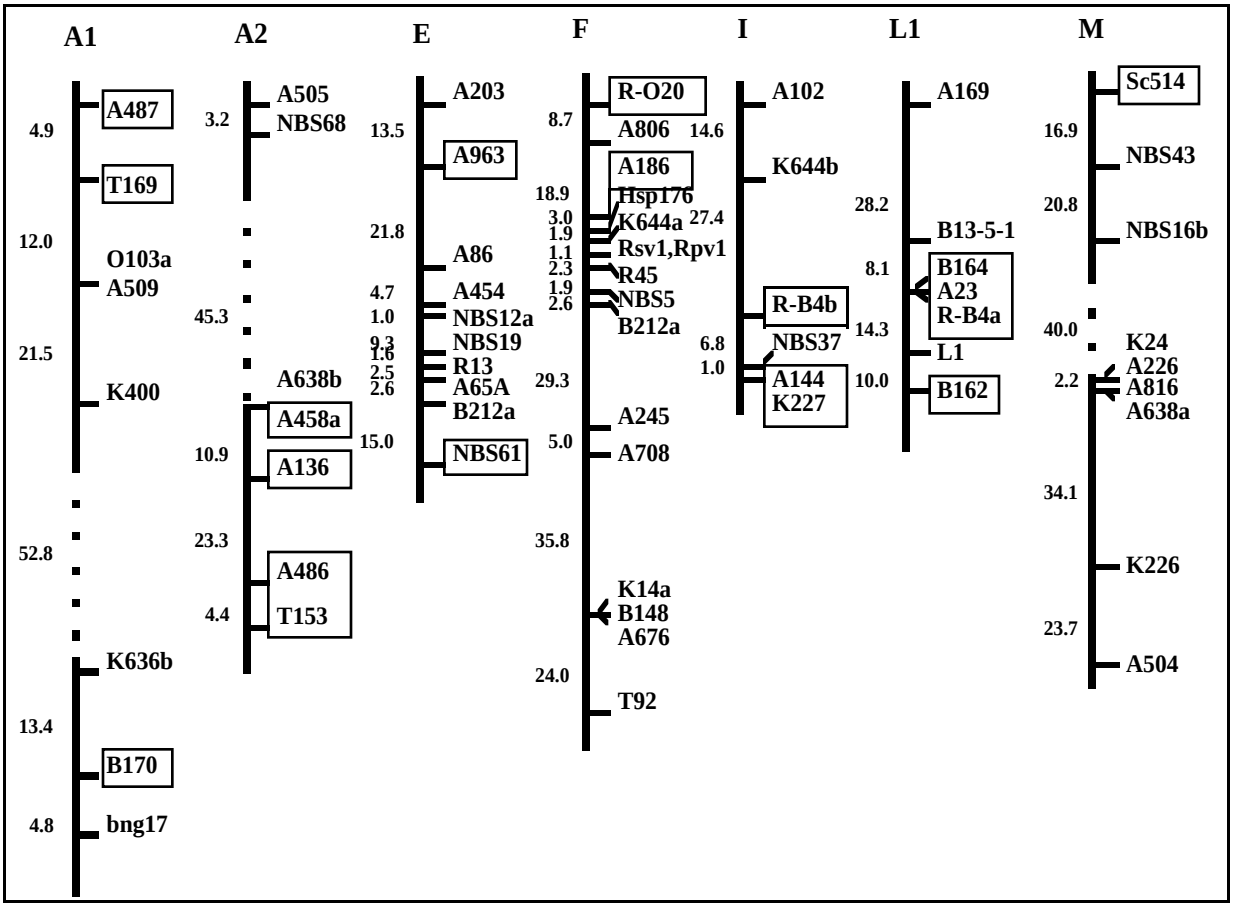


Figure 5. Seven linkage groups (adapted from Maughan et al., unpublished) showing the 19 marker loci (in boxes) associated with seed sucrose content in V71-370 x PI407162 population.

REFERENCES

Akkaya, M. S., A. A. Bhagwat, and P. B. Cregan. 1992. Length polymorhisms of simple sequence repeat DNA in soybean. Genetics 132:1131-1139.

Akkaya, M. S., R. C. Shoemaker, J. E. Specht, A. A. Bhagwat, and P. B. Cregan. 1995. Integration of simple sequence repeat DNA markers into a soybean linkage map. Crop Sci. 35:1439-1445.

Apuya, N. R., B. L. Frazier, P. Keim, E. J. Roth, and K. G. Lark. 1988. Restriction fragment length polymorphisms as genetic markers in soybean *Glycine max* (L.) Merrill. Theor. Appl. Genet. 75:889-901.

Botstein, D., R. L. White, M. Skolnick, and R. W. Davis. 1980. Construction of a genetic linkage map in human using restriction fragment length polymorphism. Am. J. Hum. Genet. 32:314-331.

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.

Concibido, V. C., R. L. Denny, S. R. Boutin, R. Hautea, J. H. Orf, and N. D. Young. 1994. DNA marker analysis of loci underlying resistance to soybean cyst nematode (*Heterodera glycines* Ichinohe). Crop Sci. 34:240-246.

Concibido, V. C., R. L. Denny, D. A. Lange, J. H. Orf, and N. D. Young. 1996. RFLP mapping and marker-assisted selection of cyst nematode resistance in PI209332. Crop Sci. 36:1643-1650.

Delannay, X., D. Rodgers, and R. Palmer. 1983. Relative genetic contribution among ancestral lines to North American soybean cultivars. Crop Sci. 23:621-624.

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.

Fehr, W. R. 1987. Principles of cultivar development. Macmillan, New York, pp. 388-400.

Griffis, G. and L. Wiedermann.1990. Marketing food-quality soybeans in Japan. American Soybean Association, P.O. Box 27300, St. Louis, Mo. 63141.

Grodzicker, T., J. Williams, P. Sharp, and J. Sambrook. 1974. Physical mapping of temperature-sensitive mutations of adenoviruses. Cold Spring Harbor Symp. Quant. Biol. 39:439-446.

Helentjaris, T., 1987. A genetic linkage map for maize based on RFLPs. Trends In Genet. 3:217-221.

Hnetkovsky, N., S. C. Chang, T. W. Doubler, P. T. Gibson, and D. A. Lightfoot. 1996. Genetic mapping of loci underlying field resistance to soybean sudden death syndrome (SDS). Crop Sci. 36:393-400.

Hsu, S. H., H. H. Hadley, and T. Hymowitz. 1973. Changes in carbohydrate contents of germinating soybean seeds. Crop Sci. 13:407-410.

Huen, M., A. E. Kennedy, J. A. Anderson, N. L. V. Lapitan, M. E. Sorrels, and S. D. Tanksley. 1991. Construction of a restriction fragment length polymorphism map for barley (*Hordeum vulgare*). Genome 34:437-447.

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

Hymowitz, T., F. I. Collins, J. Panczner, and W. M. Walker. 1972a. Relationship between the content of oil, protein, and sugar in soybean seed. Agron J. 64:613-616.

Hymowitz, T., W. M. Walker, F. I. Collins, and J. Panczner. 1972b. Stability of sugar content in soybean traits. Commun. Soil Sci. Plant Anal. 3:367-373.

Kawamura, S., 1967. Quantitative paper chromatography of sugars of the cotyledon, hull, and hypocotyl of soybeans of selected varieties. Tech. Bull. Fac. Kagawa Univ. 15:117-131.

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.

Keim, P., W. D. Beavis, J. M. Schupp. B. M. Baltazar, L. Mansur, R. E. Freestone, M. Vahedian, and D. M. Webb. 1994. RFLP analysis of soybean breeding populations: I. Genetic structure differences due to inbreeding methods. Crop Sci. 34:55-61.

Keim, P., J. M. Schupp, S. E. Travis, K. Clayton, T. Zhu, L. Shi, A. Ferreira, and D. M. Webb. 1997. A high density soybean genetic map based on AFLP markers. Crop Sci. 37:537-543.

Kiss, G. B., P. Kalo, K. Felfoldi, K. Kiss, and G. Endre. 1997. Genetic linkage map of alfalfa (*Medicago sativa*) and its use to map seed protein genes as well as genes involved in leaf morphogenesis and symbiotic nitrogen fixation. In: Biological fixation of nitrogen for ecology and sustainable agriculture. Springer. 1997. pp:279-282.

Kleinhofs, A., A. Kilian, M. A. Saghai Maroof, R. M. Biyashev, P. Hayes, F. Q. Chen, N. Lapitan, A. Fenwick, T. K. Blake, V. Kanazin, E. Ananiev, L. Dahleen, D. Kudrna, J. Bollinger, S. J. Knapp, B. Liu, M. Sorrels, M. Heun, J. D. Franckowiak, D. Hoffman, R. Skadsen, and B. J. Steffenson. 1993. A molecular, isozyme and morphological map of the berley (*Hordeum vulgare*) genome. Theor. Appl. Genet. 86:705-712.

Lark, K. G., J. M. Weisemann, B. F. Matthews, R. Palmer, K. Chase, and T. Macalma. 1993. A genetic map of soybean (*Glycine max* L.) using an intraspecific cross of two cultivars: 'Minsoy' and 'Noir 1'. Theor. Appl. Genet. 86:901-906.

Lark, K. G., J. Orf, and L. M. Mansur. 1994. Epistatic expression of quantitative tarit loci (QTL) in soybean [*Glycine max* (L.) Merr.] determined by QTL association with RFLP data. Theor. Appl. Genet. 88:486-489.

Lee, S. H., M. A. Bailey, M. A. R. Mian, T. E. Carter Jr., D. A. Ashley, R. S. Hussey, W. A. Parrott, and H. R. Boerma. 1996a. Molecular markers associated with soybean plant height, lodging, and maturity across locations. Crop Sci. 36: 728-735.

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. 1996b. 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., 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. 1996c. RFLP loci associated with soybean seed protein and oil content across populations and locations. Theor. Appl. Genet. 93:649-657.

Lin, J., J. Kuo, J. Ma, J. A. Saunders, H. S. Beard, M. H. MacDonald, W. Kenworthy, G. N. Ude, and B. F. Matthews. 1996. Identification of molecular markers in soybean comparing RFLP, RAPD and AFLP DNA mapping techniques. Plant Molec. Bio. Rep. 14:156-169.

Mansur, L. M., J. Orf, and K. G. Lark. 1993a. 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. Orf, and K. G. Lark. 1993b. 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., 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., 1994. Molecular marker analysis of quantitative trait loci influencing seed quality characteristics in soybean (*Glycine max*. (L.) Merr.). PhD. Dissertation, Virginia Polytechnic Institute and State University, Blacksburg, Virginia.

Maughan, P. J., M. A. Saghai Maroof, and G. R. Buss. 1995. Microsatellite and amplified sequence length polymorphisms in cultivated and wild soybean. Genome 38:715-723.

Maughan, P. J., M. A. Saghai Maroof and G. R. Buss. 1996a. Amplified fragment length polymorphism (AFLP) in soybean: species diversity, inheritance, and near-isogenic line analysis. Theor. Appl. Genet. 93:392-401.

Maughan, P. J., M. A. Saghai Maroof, and G. R. Buss. 1996b. 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.

McCouch, S. R., G. Kochert, Z. Y. Yu, G. S. Khush, W. R. Coffman, and S. D. Tanksley. 1988. Molecular mapping of rice chromosomes. Theor. Appl. Genet. 76:815-829.

Mian, M. A. R., M. A. Bailey, D. A. Ashley, R. Wells, T. E. Carter Jr., W. A. Parrott, and H. R. Boerma. 1996. Molecular markers associated with water use efficiency and leaf ash in soybean. Crop Sci. 36:1252-1257.

O'Donoughue, L. S., L. S. Kianian, S. F. Rayapati and P. J. Penner. 1995. A molecular linkage map of cultivated oat. Genome 38:368-380.

Openshaw, S. J., and H. H. Hadley. 1978. Maternal effects on sugar content in soybean seeds. Crop Sci. 18:581-584.

Rafalski, A. and S. Tingey. 1993. RFLP map of soybean (*Glycine max*) 2N=40. In: O'Brien S. J. (ed.) Genetic maps: Locus maps of complex genomes. Cold Spring Harbor Laboratory Press, New York, pp. 6.149-156.

Ragab, R. A., S. Dronavalli, M. A. Saghai Maroof, and Y. G. Yu. 1994. Construction of Sorghum RFLP linkage map using sorghum and maize DNA probes. Genome 37:590-594.

Ragot, M. and D. A. Hoisington. 1993. Molecular markers for plant breeding: comparison of RFLP and RAPD genotyping costs. Theor. Appl. Genet. 86:975-984.

Saghai Maroof M. A., K. M. Soliman, R. A., Jorgensen, and R. W. Allard. 1984. Ribosomal DNA spacer-length polymorphisms in barley: Mendelian inheritance, chromosomal location, and population dynamics. Proc. Natl. Acad. Sci. USA 81:8014-8018.

Sax, K., 1923. The association of size difference with seed-coat pattern and pigmentation in *Phaseolus vulgaris*. Genetics 8:552-560.

Sciumbato, G. L., and D. L. Turnage. 1992. Southern United States soybean disease estimate for 1991. pp.18-22. In G. R. Bowers, Jr. (ed.) Proc. South. Soybean Disease Workers, St. Louis, Missouri. 19-20 Feb. 1992. Texas A&M University Agric. Res. & Ext. Ctr., Beaumont, TX.

Shoemaker, R. C., R. D. Guffy, L. L. Lorenzen, J. E. Specht. 1992. Moleculer genetic mapping of soybean: Map utilization. Crop Sci. 32:1091-1098.

Shoemaker, R. C. and T. C. Olson. 1993. Molecular linkage of soybean (*Glycine max* L. Merr.). In: O'Brien S. J. (ed.) Genetic maps: Locus maps of complex genomes. Cold Spring Harbor Laboratory Press, New York, pp. 6.131-6.138.

Skorupska, H. T., R. C. Shoemaker, A. Warner, E. R. Shipe, and W. C. Bridges. 1993. Restriction fragment length polymorphism in soybean germplasm of the southern USA. Crop Sci. 33:1169-1176.

Smith, K. J. and W. Huyser. 1987. World distribution and significance of soybean in J. R. Wilcox (ed.) Soybeans: Improvement and Uses. Agronomy 16: 1-22.

Southern, E., 1975. Detection of specific sequences among DNA fragments separated by gel electrophoresis. J. Mol. Biol. 98:503-517.

Specht, J. E. and G. L. Graef. 1996. Limitations and potentials of genetic manipulation of soybean. In: Verma D. P. S. and R. C. Shoemaker (eds.) Soybean: Genetics, molecular biology and biotechnology. CAB International, Wallingford, UK, pp. 91-106.

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

Tanksley, S. D., M. W. Ganal, J. P. Prince, M. C. deVincente, M. W. Bonierbale, P. Brown, T. M. Fulton, J. J. Giovanonni, S. Grandillo, G. B. Martin, R. Messeguer, J. C. Miller, L. Miller, A. H. Peterson, O. Pineda, M. Roder, R. A. Wing, W. Wu, and N. D. Young. 1992. High density molecular linkage maps of the tomato and potato genomes: biological inferences and practical applications. Genetics 132:1141-1160.

Tautz, D., 1989. Hypervariability of simple sequences as a general source of polymorphic DNA markers. Nucleic Acids Res. 17:6463-6471.

Vos, P., R. Hogers, M. Bleeker, M. Reijans, T. van de Lee, M. Hornes, A. Frijters, J. Pot, J. Peleman, M. Kuiper, and M. Zabeau. 1995. AFLP: a new technique for DNA fingerprinting. Nucleic Acids Res. 23:4407-4414.

Webb, D. M., B. M. Baltazar, A. P. Rao-Arelli, J. Schupp, K. Clayton, P. Keim, and W. D. Beavis. 1995. Genetic mapping of soybean cyst nematode race-3 resistance loci in the soybean PI 437.654. Theor. Appl. Genet. 91:574-581.

Williams, J. G. K., A. R. Kubelik, K. J. Livak, J. A. Rafalski, and S. V. Tingey. 1990. DNA polymorphisms amplified by arbitrary primers are useful as genetic markers. Nucleic Acids Res. 18:6531-6535.

Yu, Y. G., M. A. Saghai Maroof, G. R. Buss, P. J. Maughan, and S. A. Tolin. 1994. RFLP and microsatellite mapping of a gene for soybean mosaic virus resistance. Phytopathology 84:60-64.

Yu, Y. G., M. A. Saghai Maroof, and G. R. Buss. 1996. Divergence and allelomorphic relationship of a soybean virus resistance gene based on tightly linked DNA microsatellite and RFLP markers. Theor. Appl. Genet. 92:64-69.

Yu, Y. G., G. R. Buss, and M. A. Saghai Maroof. 1996. Isolation of superfamily of candidate disease-resistance genes in soybean based on a conserved nucleotide binding site. Proc. Natl. Acad. Sci. U.S.A. 93:11751-11756.

VITA

Mine Cicek was born January 19, 1970 in Istanbul, Turkiye. Mrs. Cicek attended Uskudar American Academy for Girls highschool in Istanbul, Turkiye, where she had seven years of middle school and highschool education and graduated in May 1988. She then attended Hacettepe University, Ankara, Turkiye. She completed her studies in the Faculty of Science, Department of Biology and received her Bachelor of Science degree in May 1992. She got married to Muzaffer Cicek in 1993 and came to Virginia Tech, Blacksburg, VA, United States. In 1994, she worked at Corning plant in Christiansburg, VA as an assistant in a laser project team. She decided to continue her graduate education in plant breeding and molecular genetics. She accepted a graduate research assistantship in the Department of Crop and Soil Environmental Sciences in 1995. She has a two months old son, Hakan Mirzahan Cicek.