SOYBEAN QTL FOR YIELD AND YIELD COMPONENTS ASSOCIATED WITH GLYCINE SOJA ALLELES

Dandan Li
University of Kentucky, dli2@uky.edu

Recommended Citation
https://uknowledge.uky.edu/gradschool_diss/331

This Dissertation is brought to you for free and open access by the Graduate School at UKnowledge. It has been accepted for inclusion in University of Kentucky Doctoral Dissertations by an authorized administrator of UKnowledge. For more information, please contact UKnowledge@lsv.uky.edu.
ABSTRACT OF DISSERTATION

Dandan Li

The Graduate School
University of Kentucky
2006
SOYBEAN QTL FOR YIELD AND YIELD COMPONENTS ASSOCIATED WITH GLYCINE SOJA ALLELES

ABSTRACT OF DISSERTATION

A dissertation submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy in the College of Agriculture at the University of Kentucky

By

Dandan Li

Lexington, Kentucky

Director: Dr. Todd W. Pfeiffer, Professor of Crop Science

Lexington, Kentucky

2006

Copyright © Dandan Li 2006
USA soybean germplasm has a narrow genetic base that could be augmented by alleles from the wild species *Glycine soja* which positively influence agronomic traits. The objective of this study was to identify such alleles for yield and yield component QTL (quantitative trait loci). Two populations of 150 BC\textsubscript{2}F\textsubscript{4} lines were generated from a mating between recurrent parent *Glycine max* ‘7499’ and donor parent *Glycine soja* PI 245331 with one line in each population tracing back to the same BC\textsubscript{2} plant. Population A was used for the QTL identification analysis and population B was used for the QTL verification test. The population A lines were genotyped at 120 SSR marker loci and one phenotype marker, covering a total map length of 1506 cM in 20 linkage groups with an average interval size of 12.5 cM. There were nine putative QTL significantly (P<0.0001, LOD>3.0) associated with yield and yield component traits across 3 environments. One QTL for seed yield was identified using the combined data; the *G. soja* allele at satt511 on LG-A1 was associated with increased seed yield (LOD=4.3) with an additive yield effect of 190 – 235 kg ha\textsuperscript{-1} depending on the QTL analysis method. The phenotypic variance accounted for by the QTL at satt511 was 12%. This QTL also provided a significant yield increase across environments in the validation population; lines that were homozygous for the *G. soja* allele at satt511 demonstrated a 6.3% (P=0.037) yield increase over lines that were homozygous for the *G. max* allele. One seed filling period QTL was identified at satt335 (LOD=4.0) on LG-F with an additive effect of +1 day. This QTL also provided a +1 day additive effect (LOD=3.3) on maturity. These results demonstrate the potential of using exotic germplasm to improve soybean yield.

**KEYWORDS:** QTL, *Glycine max*, *Glycine soja*, Genetic Diversity, Yield Improvement

Dandan Li
May 15\textsuperscript{th}, 2006
SOYBEAN QTL FOR YIELD AND YIELD COMPONENTS ASSOCIATED WITH GLYCINE SOJA ALLELES

By

Dandan Li

Todd M. Pfeiffer
Director of Dissertation

Charles T. Dougherty
Director of Graduate Studies

May 15th, 2006
RULES FOR THE USE OF DISSERTATIONS

Unpublished dissertations submitted for the Doctor’s degree and deposited in the University of Kentucky Library are as a rule open for inspection, but are used only with due regard to the rights of the authors. Bibliographical references may be noted, but quotations or summaries of parts may be published only with permission of the author, and with the usual scholarly acknowledgments.

Extensive copying or publication of the dissertation in whole or in part also requires the consent of the Dean of the Graduate School of the University of Kentucky.

A library that borrows this dissertation for use by its patrons is expected to secure the signature of each user.
SOYBEAN QTL FOR YIELD AND YIELD COMPONENTS ASSOCIATED WITH GLYCINE SOJA ALLELES

DISSERTATION

A dissertation submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy in the College of Agriculture at the University of Kentucky

By

Dandan Li

Lexington, Kentucky

Director: Dr. Todd M. Pfeiffer, Professor of Crop Science

Lexington, Kentucky

2006

Copyright © Dandan Li 2006
ACKNOWLEDGMENTS

I would like to express my great gratitude and sincere appreciation to Dr. Todd Pfeiffer for his invaluable contribution and direction.

His academic guidance, dedication, and extreme patience made possible the successful completion of this dissertation. His knowledge, insightful comments and high standards were instrumental in helping develop my research and improving my writing skills. Next, I wish to express my deepest appreciation to the members of my dissertation committee, Dr. Dave A. VanSanford, Dr. Dennis B. Egli, Dr. Paul L. Cornelius, and Dr. Haining Zhu for their invaluable comments, critiques, and advice that enriched this dissertation.

I would like to give my sincere thanks to Dave L. Pilcher, Tom R. Stefaniak, Carrie A. Knott; their full support and friendship made my life happier and guaranteed my research going smoothly and successfully.

Finally, my thanks and appreciation to my husband Shengming Yang for his love, understanding, and tireless support during those busy days! Under God’s mercy and grace, we are studying in the same department, enjoying his love together. All the honors and praise belong to God, our savior!
# TABLE OF CONTENTS

ACKNOWLEDGMENTS.................................................................................................iii

LIST OF TABLES........................................................................................................vi

LIST OF FIGURES.......................................................................................................vii

LIST OF FILES...........................................................................................................viii

CHAPTER ONE

INTRODUCTION........................................................................................................1

CHAPTER TWO

LITERATURE REVIEW............................................................................................4

Yield Improvement in Soybean.................................................................................4
  *U.S. Soybean Yield Increase*...............................................................................4
  *International Soybean Yield Increase*.................................................................4

Gene pools of Soybean............................................................................................4
  *Soybean GP-1*.......................................................................................................5
  *Soybean GP-2*.......................................................................................................5
  *Soybean GP-3*.......................................................................................................5

Soybean Gerplasm Enhancement and Genetic Diversity........................................6
  *Conventional Plant Breeding*...............................................................................6
  *Favorable Allele Introgression*............................................................................8
  *Marker Analyses*................................................................................................10

QTL Mapping..........................................................................................................13
  *Mapping Method*................................................................................................14
  *Advanced Backcross*..........................................................................................15
  *Soybean Molecular Marker Map And Yield QTL*...............................................17

Why Look For Seed Filling Period QTL.................................................................21

CHAPTER THREE

Thesis Research Objectives....................................................................................27

CHAPTER FOUR

MATERIAL AND METHODS...................................................................................28

  *Parental Germplasm and Population Development*..........................................28
  *Phenotypic Trait Measurement*..........................................................................29
  *Statistical Analysis*.............................................................................................30
  *Leaf Collection And DNA Isolation Procedures*.................................................31
  *Phenotypic Marker*............................................................................................31
CHAPTER FIVE

RESULTS ..................................................................................................................45
Polymorphism and Marker Segregation .........................................................45
Linkage Map .................................................................................................45
Field Data Analysis ......................................................................................46
Trait Correlations .........................................................................................48
QTL Discovery .............................................................................................48
  Seed Yield .............................................................................................49
  100-Seed Weight ..................................................................................50
  Seed Filling Period ...............................................................................51
  Maturity .................................................................................................52
  Lodging ....................................................................................................53
  Height ........................................................................................................54
QTL Confirmation ..........................................................................................54

CHAPTER SIX

DISCUSSION .............................................................................................................78
Why PI 245331 .................................................................................................78
Power of the QTL Test .................................................................................78
Mapping Population ......................................................................................80
Seed Yield QTL .............................................................................................81
Seed Filling Period QTL .................................................................................83
Maturity QTL .................................................................................................84
QTL Association ...........................................................................................84
Future Perspectives .........................................................................................84

REFERENCES .........................................................................................................89

VITA ......................................................................................................................96
**LIST OF TABLES**

Table 1. Development of BC$_2$F$_4$ Population ........................................36
Table 2. $X^2$ Values for Each Marker, Individual Linkage Group, and all Markers ....57
Table 3. Pearson Correlation Coefficients for Yield Between Environmental Pairs .....58
Table 4. Descriptive Statistics for G. Max and BC$_2$F$_4$ Population ..................59
Table 5. Analysis of Variance on Each Trait in Population A at Each Environment and for the Combined Data ........................................................................60
Table 6. Pearson Correlation Coefficients Between Seed Yield and Other Traits Within Different Maturity Sets .................................................................61
Table 7. Pearson Correlation Coefficients for Trait Pairs ....................................62
Table 8. QTLs Identified By MR for Combined Environments ...............................63
Table 9. QTLs Identified By SIM for Combined Environments ..............................64
Table 10. QTLs Identified By CIM for Combined Environments ............................65
Table 11. QTLs Identified By MR for Each Environments .....................................66
Table 12. QTLs Identified By SIM for Each Environments ....................................68
Table 13. Allelic Means on Seed Yield, SFP, and Maturity for Population A and B ....70
Table 14. Allelic Means and Class Difference for The Two Yield QTL ................71
Table 15. Analysis of Variance for Population B at Each Environment ................72
Table 16. Analysis of Variance for Population B at Combined Environment ........74
Table 17. Mean Squares and Heritability for All the Traits ................................87
Table 18. Yield QTL identified from unimproved accessions ..............................88
LIST OF FIGURES

Figure 1. Graphic Representation of Soybean Yield……………………………………23
Figure 2. Genetic Linkage Map Constructed for Population A…………………………37
Figure 3  Graphic Depiction of the LOD Score and Additive Effects.......................75
<table>
<thead>
<tr>
<th>File Name</th>
<th>Size</th>
</tr>
</thead>
<tbody>
<tr>
<td>DissertationDandanLi.pdf</td>
<td>2.43MB</td>
</tr>
</tbody>
</table>
CHAPTER ONE
INTRODUCTION

The soybean (Glycine max (L.) Merr.), rich in seed protein and oil, is an economically important leguminous seed crop for feed, oil, and soy food products. In the international trade markets soybean is ranked number one in world oil production (48%) among the major crops, such as cottonseed (Gossypium hirsutum, barbadense), peanut (Elaeis guineensis), sunflower seed (Helianthus annuus), rapeseed (brassica napus), coconut (Cocos nucifera), and palm kernel (Elaeis guineensis). Presently, soybean is a world crop, cultivated widely in the United States, Brazil, Argentina, China, and India. The United States is the leader in soybean production.

Soybean was domesticated from the wild soybean, Glycine soja Sieb. & Zucc. It is likely that strong selection pressure exerted by humans on the genetic diversity found in the wild resulted in rapid and radical changes in soybean. Certain traits, such as non-shattering of seeds and compact growth habit, would have been selected by early agriculturists. Selective propagation of lines containing these favorable mutations would have resulted in a progressive narrowing of the genetic base of subsequent populations.

Following domestication, the genetic variation in crop plants has been continually reduced by another force—modern plant breeding. Successful soybean breeding efforts in North America have resulted in the release of more than 300 publicly developed cultivars in the past 50 yr. To develop new cultivars, breeders generally use current high-yielding cultivars as parents to create populations from which to make selection. One negative consequence of this breeding methodology is that new cultivars become more closely related after each cycle. Using closely related cultivars as parents to create populations can reduce genetic diversity and the subsequent gains made from selection (Manjarrez et al., 1997; Kisha et al., 1997).

Soybean is a good example of a crop with a very narrow genetic base. The North American soybean gene pool can be traced back to only 50 plant introductions with more than 80% of the northern and southern gene pool originating from fewer than a dozen introductions (Delannay et al. 1983; Gizlice et al., 1994). The limited genetic diversity of crops renders them more vulnerable to insect and disease epidemics and jeopardizes the potential for sustained genetic improvement over the long term. This risk was brought sharply into focus in 1970 with the outbreak of southern corn leaf blight. This disease drastically reduced corn yields in the
United States. The cause of this dramatic loss was attributed to extensive use of a single genetic male sterility factor that, unfortunately, was genetically linked to disease susceptibility.

The traditional approach to the utilization of exotic germplasm is to screen entries from a gene bank for a clearly defined character recognizable in the phenotype. Once a line with the desired characteristic has been found, it is crossed with an elite cultivar in order to introduce the genes from the exotic donor into the cultivated type. This approach works well when the trait of interest is controlled by one or a few genes. For example, wild germplasm has been used with great success in breeding for simply inherited resistance to diseases and insects (Concibido and Denny, 1994; Webb and Baltazar, 1995; Concibido and Denny, 1997; Tamulonis and Luzzi, 1997; Rector and All, 2000; Terry and Chase, 2000; Klos and Paz, 2000; Hayes and Ma, 2000).

Most traits important to agriculture, such as yield, are controlled by polygenes. Crop domestication and plant breeding have been extremely successful at increasing the frequency of beneficial alleles for yield at many loci. But undoubtedly, many beneficial alleles have been left behind because of the bottlenecks imposed by domestication coupled with years of modern breeding and selection within adapted gene pools. It has been shown that wild species may have one or more alleles of genes with a positive influence on one or more agronomic traits ( Tanksley and McCouch, 1997). Such favorable alleles might be beneficial if introduced into elite cultivars lacking such alleles.

Molecular markers have provided a powerful new tool for breeders to search for new sources of variation and to investigate genetic factors controlling quantitatively inherited traits. Genetic linkage maps based on molecular markers have now been developed for most major crop species. These maps have been important in a number of applications. But the most profound impact has been in the study of complex and quantitatively inherited traits, such as yield. Molecular linkage maps have made it possible to identify, map, and study the effects of the individual loci that control a quantitatively inherited trait (quantitative trait loci, or QTL).

Relative to a multi-genetic quantitatively inherited trait (e.g. yield), the identification of favorable alleles is difficult, because many more undesirable and deleterious alleles will also segregate in populations derived from $G. \text{max} \times G. \text{soja}$ matings. The use of advanced backcross populations has been proposed as a means to solve this problem ( Tanksley and Nelson, 1996). Backcrossing reduces the number of donor-derived wild species alleles present in any given backcross-derived line, thus allowing the many undesirable or deleterious alleles to be readily
exposed, but more importantly, allowing the few favorable donor alleles to be more easily recognized. Those few favorable alleles, now present in a more elite background, can then be easily introgressed into the elite germplasm pool with minimal concurrent introgression of any unfavorable alleles.

The strategy involves the use of molecular linkage maps and a breeding technique, referred to as the advanced backcross QTL method, which allows a subset of alleles from the wild or exotic plant to be examined in the genetic background of an elite cultivar. The molecular linkage map is used to identify the chromosomal position of “wild” alleles that have been transmitted into the progeny, to determine which of the wild species introgressions are associated with the superior performance of lines, and to purify lines so that they contain only a specific “wild QTL” in an elite genetic background.

Using markers to identify QTL, this research was conducted to test the yield differences among 147 BC₂F₄ lines derived from the cross *G. max* × *G. soja*. A split plot experiment repeated over two locations was used. The 147 lines were separated into three different maturity sets based on their maturity date. Each maturity set had 2 replications per environment. The null hypothesis in this research is that the yields of all soybean genotypes tested are the same. The research hypothesis is that yield differences exist in soybean BC₂ genotypes within each environment. The effects of environment, whole plot (maturity sets) and subplot (lines/maturity set) and their interaction effects were tested. At the same time, marker analysis data were collected. Based on the field data and lab data the yield QTL maps were built.
CHAPTER TWO
LITERATURE REVIEW
Yield Improvement in Soybean

U.S. Soybean Yield Increase

Improvement of elite soybean germplasm in North America may have been limited by genetic bottlenecks during its development. Reports show that yield improvements in soybean continue at a modest rate of approximately 30 kg ha\(^{-1}\) year\(^{-1}\) or about 1% year\(^{-1}\) (Specht et al. 1999; Wilcox 2001).

Specht et al. (1999) reported U.S. soybean yield had increased at a linear rate of 22.6±0.7 kg ha\(^{-1}\) yr\(^{-1}\) from 1924 to 1998. However during the last quarter century (1972 to 1998) soybean yield improvement has been 40% faster at a rate of 31.2±4.8 kg ha\(^{-1}\) yr\(^{-1}\). This yield improvement was due to the adoption of agricultural innovation in the form of genetic, agronomic, or genetic × agronomic technologies, including higher yielding cultivars and management practices.

Wilcox (2001) subsequently reported public soybean yield improvement in the northern USA by maturity group (MG), showing that over the last 60 years, rates of yield improvement in kg ha\(^{-1}\) yr\(^{-1}\) were 21.6 (MG 00), 25.8 (MG 0), 30.4 (MG I), 29.3 (MG II), 30.6 (MG III), and 29.5 (MG IV).

International Soybean Yield Increase

Singh and Hymowitz (1999) noted that the limited genetic diversity may be one reason why North American public and proprietary soybean cultivars have made only modest advances in yield for the past 20 years (Figure 1). A similar situation prevails for soybean yields in India. The genetic base of soybean cultivars grown in India is extremely narrow. Soybean cultivars are direct introductions from the United States, selections from the introduced germplasm, and pure lines derived from single-cross populations (Singh and Hymowitz, 1999). Unlike India, Chinese soybean cultivars have a high level of genetic diversity (Cui et al. 2000). In 1961, soybean yield in China was 626 kg/ha, increasing to 1893 kg/ha in 2002 (Figure 1).

Gene Pools of Soybean

Soybean was domesticated from the wild soybean, *Glycine soja* Sieb. & Zucc. (Formerly *Glycine ussuriensis* Regal & Maack) (Fukuda, 1933). Wild soybean is an annual weedy-form climber, whose pods contain black seeds that shatter at maturity. The plant grows wild in China
and adjacent regions of Russia, Korea, Taiwan and Japan (Hymowitz 1970). The wild soybean is rich in seed protein (31.1-52.4%) but low in oil (9.1-11.9%) content (Hymowitz, 1972). The cultivated soybean and its progenitor \textit{G. soja} belong to the subgenus \textit{Soja} (Moench.) F.J. Herm., and both are cross compatible, contain \(2n=40\) chromosomes, and produce vigorous fertile intermediate \(F_1\) hybrids (Singh et al. 1999).

Harlan and de Wet (1971) proposed the concept of three gene pools, primary (GP-1), secondary (GP-2), and tertiary (GP-3), based on the success rate of hybridization among species.

\textbf{Soybean GP-1}

GP-1 consists of biological species for which crossing within the gene pool is easy. Crosses within this pool result in hybrids that are vigorous, exhibit normal meiotic chromosome pairing, and possess total seed fertility. Gene segregation in these crosses is normal, and gene exchange is generally easy. GP-1 was further subdivided into subspecies A, which includes cultivated races, and subspecies B, which includes spontaneous races. Soybean (\textit{G. max}) cultivars and landraces and their wild annual progenitor, \textit{G. soja}, are included in GP-1 (Singh and Hymowitz, 1999).

\textbf{Soybean GP-2}

Harlan and de Wet (1971) defined GP-2 as, “all species that can be crossed with GP-1 with at least some fertility in \(F_1\).” A GP-2 for soybean (\textit{G. max}) has not been identified.

\textbf{Soybean GP-3}

GP-3 is the extreme outer limit of potential genetic resource. Hybrids between GP-1 and GP-3 are anomalous, lethal, or completely sterile, and gene transfer is not possible or requires radical techniques (Harlan and de Wet, 1971). Based on the definition, GP-3 includes 16 wild perennial species of the subgenus \textit{Glycine}. These species are indigenous to Australia and are geographically isolated from \textit{G. max} and \textit{G. soja} (native of China). Extensive plant exploration trips by T. Hymowitz (1984) and A.H.D. Brown in Australia and to the South and West Central Pacific Islands have enriched the collection of wild perennial \textit{Glycine} species germplasm. A total of 1025 accessions are being maintained at the University of Illinois (Singh and Hymowitz, 1999).

The soybean germplasm is a rich reservoir, with more than 170000 \textit{G. max} accessions maintained by more than 160 institutions in nearly 70 countries (Carter et al., 2004). There are fewer than 10000 \textit{G. soja} accessions, and about 3535 accessions of wild perennial \textit{Glycine}
species in germplasm banks throughout the world (Carter et al., 2004). Carter et al. (2004) reported a total of only 18114 accessions of *G. max* and 1114 *G. soja* accessions in the USDA soybean data base. The USDA perennial *Glycine* collection grew from six species in 1975 to a current 13 species and approximately 1000 accessions in 2001.

**Soybean Germplasm Enhancement and Genetic Diversity**

The soybean is an economically important leguminous crop for feed, oil, and soyfood products. Despite its economic importance, the genetic base of soybean cultivars is extremely narrow. A germplasm source which is collected but never used contributes as little to plant breeding progress as germplasm which is never collected. In the past our germplasm collections have been used to find single gene sources of resistance to diseases and insects and tolerance to abiotic stresses. This has been true of all our major crops (Shands and Weisner 1996). But, as Tanksley and McCouch (1997) point out, our germplasm collections as reservoirs of genes for improvement of quantitative production traits such as yield have been vastly underutilized.

**Conventional Plant Breeding**

The first soybean cultivars grown in the United States were direct introductions from Asia and selections from the introduced germplasm (Singh and Hymowitz, 1999). During the last five decades, soybean breeding has been limited to hybridization within GP-1, and selection has focused on high yielding cultivars with high seed protein and oil concentrations. Delannay et al. (1983) examined 158 U.S. and Canadian soybean cultivars and noted that all of this germplasm could be traced to as few as 50 plant introductions (PIs). The authors also observed that more introductions contributed to the northern elite gene pool than to the southern elite gene pool.

Gizlice et al. (1993) examined pedigrees of North American public soybean cultivars released from 1947 to 1988 and calculated coefficients of parentage for the available 258 public soybean cultivars (northern and southern U.S.). Coefficient of parentage (COP) is the probability that randomly chosen alleles at any given locus in the two cultivars are identical by descent, and is a metric commonly used to express the relatedness of cultivars when choosing parents for mating. Mean COP for 258 public soybean cultivars released between 1947 and 1988 was 0.13. Coefficient of parentage increased linearly over time from an initial value of 0 prior to 1947 to a value of 0.15 among the 63 cultivars released between 1984 and 1988. Cultivars released after 1983 carry 50% more genes in common than cultivars released before 1954. The authors found that over time the COP in both northern and southern gene pools had increased 25%, and was
likely to continue to increase. The mean COP was 0.18 for all northern cultivars and 0.23 for the southern cultivars, indicating northern breeders were using more diverse matings than southern breeders. However, the mean COP between the northern and southern cultivars was only 0.04, suggesting that breeders have tended to avoid the hybridization of lines differing widely in maturity and have kept northern and southern gene pools identifiably distinct.

Gizlice et al. (1994) found that the genetic base of 258 public cultivars released between 1947 and 1988 was fully defined by 80 ancestors. But these ancestors’ contributions to the genetic base of soybean were unequal and ranged from 0.00001 to 12.1%. More than half of the genetic base for North American soybean is contributed by only six ancestors. These are Mandarin (Ottawa), CNS, Richland, S-100, and Lincoln’s two unknown parents. Only 17 ancestors contribute at least 1% each to the genetic base of North American cultivars. The contributions of these 17 ancestors amount to 84% of the total genetic base of soybean.

Gizlice et al. (1994) also found that the northern and southern genetic bases are equally narrow. The southern genetic base is dominated by 17 ancestors that contributed 94% of the genes in southern cultivars. Nearly half of this contribution came from CNS and S-100. Two-thirds of the southern genetic base was defined by just five ancestors: CNS, S-100, Tokyo, PI54610, and Roanoke. Ten ancestors contributed 80% of the genes in southern breeding programs. In the north, the unknown parents of Lincoln were the dominant ancestors contributing a quarter of the genes to the northern genetic base. The four next most important ancestors, Mandarin, Richland, A.K. and Mukden contributed 40% of the genetic base.

Gizlice et al. (1996) noted that the genetic relatedness of North American soybean may impede long-term breeding progress. They proposed that breeders can diversify applied programs by capitalizing upon genetic patterns that may exist in cultivated germplasm. The North-South variability is one diversity pattern that was created by applied breeding. Other factors influencing diversity that were examined included the COP within and between maturity groups and location of breeding programs. Based on the COP, the South-North distinction accounted for only 21% of variability in cultivar relationship. Maturity groups independently accounted for 32% and breeding programs 42% of the total variation in cultivar relationships. Maturity groups and breeding programs jointly explained 48% of the total variation in COP. This result indicated that individual breeding programs have a significant influence on the pattern of diversity beyond what is accounted for by maturity alone.
Cui et al. (2000) quantified the genetic diversity in 651 Chinese soybean cultivars released from 1923 to 1995 via coefficient of parentage (COP). A very low mean COP of 0.02 was found in Chinese soybean cultivars, suggesting the presence of a potentially high level of genetic diversity in Chinese soybean breeding. The geographical region of production, province of origin, intended cropping system and era of release were also studied. The cultivars from the three major soybean growing regions of China, Northeastern, Northern, and Southern China, shared little common ancestry with each other or with U.S. cultivars. Each of these three regions showed a broader genetic base than that of the U.S. cultivars. Cultivar pools from each of the three growing regions of China were almost completely unrelated to each other and exhibited low within-region mean COP values (< 0.06). Mean COP values within and between-provinces were low (0-0.2). Cropping system and release eras also exhibited low within- and between-COP relationships (all < 0.07). The low COP values for Chinese soybean cultivars resulted from Chinese breeder initiatives to introduce new germplasm into applied Chinese breeding since the 1970s and from a strong tendency to avoid the mating of related parents. Half- and full-sib matings and backcrossing are almost absent from Chinese pedigrees. Only 7% of the genes in Chinese cultivars could be traced to U.S. cultivars and breeding lines. Chinese cultivars could be used to increase diversity in U.S. cultivars.

**Favorable Allele Introgression**

Introgression of diverse germplasm into the current soybean (*G. max*) genetic base can increase genetic variability and may lead to greater gains from selection. Soybean researchers have investigated using *G. max* plant introductions (PI) as sources of favorable alleles to broaden the genetic base.

Schoener and Fehr (1979) evaluated the relationship between the percentage of PI germplasm in a progeny population and its agronomic performance. Populations AP1 to AP5 (Fehr and Clark, 1973) containing 100, 75, 50, 25, and 0% PI germplasm were evaluated. All populations possessed high-yielding lines, and yield generally increased as the percentage of PI parentage decreased.

Vello et al. (1984) evaluated genetic variability in populations AP10 to AP14 (Fehr and Cianzio, 1981) developed from 40 PIs and 40 cultivars. They found that as the percentage of PI parentage increased, yield decreased, and they concluded that PI germplasm couldn’t enhance the yield of soybean cultivars in the short term. Ininda et al. (1996) evaluated the populations
AP10 to AP14 after three cycles of recurrent selection and reported that gain from selection was greatest in the 0% PI population, AP14. They concluded that selection from populations developed with elite parents continues to be the most effective method of developing higher yielding cultivars.

Thompson and Nelson (1998) identified 57 experimental lines which contained anywhere from 25%-100% PI germplasm. These lines were derived from 13 matings involving seven U.S. elite cultivars and seven *G. max* PIs. One of their findings was that two backcross-derived lines with different PIs as donor parents had yields greater than the recurrent parent Beeson 80. The BC$_1$ line with PI 407720 as the donor parent yielded 199 kg ha$^{-1}$ higher than Beeson 80. The BC$_1$ line with PI 68658 as the donor parent had a 198 kg ha$^{-1}$ advantage over Beeson 80. This research demonstrated that some PI germplasm, when backcrossed to elite germplasm, resulted in progeny lines with better agronomic performance than the recurrent parent. The authors suggested that such lines, if used in subsequent matings, would add to the genetic diversity of the elite soybean gene pool.

Scott and Kephart (1997) evaluated the potential of crosses of adapted × introduced soybean lines for improving yield and protein and oil concentrations. Five adapted soybean cultivars were crossed in selected combinations with four PI lines to develop F$_3$-derived lines. Thirty F$_{3:5}$ lines and parents from each of eight different crosses were evaluated. They found significant genetic variation for yield only in the two adapted × PI crosses, Norsoy × PI437.666 and Kato × PI36.653. The adapted × adapted crosses had higher yield than all but one adapted × PI cross. Mean yield for Norsoy x PI297.513 (adapted × PI) was 298 g m$^{-2}$, equal to the mean yield of the adapted × adapted cross Weber × Kato. Norsoy × PI297.513 was also among the highest for oil concentration. The authors concluded that PI germplasm may be used to improve or maintain oil concentration and still make adequate yield improvement.

But the soybean germplasm collection does not consist only of *G. max* germplasm. There are also over 1000 *G. soja* accessions in the collection (Palmer et al., 1996) ranging from maturity group 000 to maturity group X and collected from eastern Russia (24%), Korea (34%), Japan (29%), and China (17%).

*Glycine soja*, the progenitor of *G. max*, may be an excellent source of genetic variability, although it harbors several undesirable genetic traits, for example, vining, lodging susceptibility, lack of complete leaf abscission, seed shattering, and small seed with black seed coat (Carpenter
and Fehr, 1986). However, *G. soja* constitutes unimproved germplasm and, during the course of selection in successive backcross generations, undesirable traits can be discarded (Singh and Hymowitz, 1999).

Ertl and Fehr (1985) developed two populations from *G. max × G. soja* matings. Five backcrosses were conducted after the matings. The F2 plants from the backcross populations were selected for the absence of the deleterious phenotypes. Progeny of selected plants from the backcross populations were evaluated for seed yield, maturity, lodging, and height in four environments during 2 years. But the results showed that the introgression of *G. soja* germplasm into the two soybean cultivars was not an effective method for increasing their yield potential.

Carpenter and Fehr (1986) mated two *G. max* with two *G. soja* lines to determine the minimum number of backcrosses needed for the resultant BC populations to perform on par with the recurrent parent (*G. max*). The percentage of segregates with an agronomic rating equal to the *G. max* parent was zero in BC0 and BC1, 2% in BC2, 22% in BC3, 51% in BC4, and 65% in BC5. From these results the authors draw the conclusions that in order to recover any segregates that are agronomically acceptable when at least 50 plants are backcrossed each generation, at least two backcrosses are required.

Suárez et al. (1991) evaluated two backcross populations to determine the phenotypic effects on yield and other traits with isozyme markers in a mating of *G. max × G. soja*. For the cross between A80-244036 (*G. max*) and PI 326581 (*G. soja*), the parents possessed different alleles at six isozyme loci. For the cross between A81-157007 (*G. max*) and PI 342618A (*G. soja*), the parents differed at eight isozyme loci. 480 BC2F4:6 lines from the two crosses were evaluated for 13 traits in two environments. In both crosses, significant differences between donor parent and recurrent parent were observed for all traits, including yield. For most of the isozyme loci, BILs (backcross inbred lines) carrying the donor parent allele were lower yielding than the recurrent parent. But, in the PI326581 mating, the *G. soja* acid phosphatase allele *Ap-c* was associated with a significant positive yield improvement of 304 kg/ha. In the 342618A mating, the *G. soja* allele at the two interacting loci diaphorase and phosphoglucomutase was observed to have a significant 342 kg/ha positive yield effect.

**Marker Analyses**

The use of wild species of soybean offers a vast genetic resource that can broaden soybean’s genetic base (Concibido et al. 2003). In the past, breeders have tried to tap into the
exotic germplasm to improve soybean and other crops with limited success, except in the case of disease resistance genes, where exotic germplasm has had a major impact. Manipulating any beneficial genes to improve yield is complex as these are often trapped in tight genetic linkages with deleterious traits like lodging, pod shattering and susceptibility to pests and diseases. Furthermore, the use of wild germplasm in a breeding program requires several cycles of backcrossing to the adapted parent to recover most of the desirable agronomic traits, which can be costly, time-consuming, and is difficult to achieve while retaining the targeted allele from the exotic donor. Molecular markers have provided a powerful new tool for breeders to search for new sources of genetic variation for soybean grown in U.S and around the world.

Kisha et al. (1998) observed the diversity among 165 lines from five gene pools consisting of ancestral lines, elite lines from northern or southern U.S. production areas, and PIs with a northern or southern maturity. The relationship among the pools was evaluated by analysis of 53 RFLPs (Restriction Fragment Length Polymorphism) and by coefficient of parentage. The average simple matching coefficient (SMC), which is defined as the probability of sampling genes from two lines that are identical by state, was calculated between all pairs of genotypes in this study. SMC is analogous to CP, which is the probability of sampling genes that are identical by descent. In comparisons of the SMC, the between pool average SMC was significantly less than within pool values. This indicates that the between pool lines are more diverse than the within pool lines. Cluster analysis was conducted on both the CP and SMC values. The clustering results indicated that much of the RFLP diversity among elite lines was predicted by their pedigree of a few major ancestors.

Lorenzen et al. (1995) conducted an RFLP marker allele frequency analysis of 64 cultivars, which included 28 successful proprietary cultivars released between 1939 and 1979 and 36 other cultivars released in the 1980s and early 1990s. The 64 genotypes were screened with 198 RFLP probes capable of detecting 217 loci. Fifty six percent of the 217 tested loci were monomorphic among all 64 genotypes. In contrast, all of these 217 loci were polymorphic in the G. max × G. soja mapping population (Shoemaker and Olson, 1993). Lorenzen et al. (1995) speculated that the loci that were monomorphic in their elite soybean lines represent genomic regions where G. max and G. soja have diverged.

Sneller et al. (1997) evaluated the association of agronomic value and genetic diversity of selected PIs and conducted a preliminary comparison of their value in the South with that of
selected northern U.S. cultivars. Thirty one PIs, 15 elite southern and 11 northern cultivars were planted in five environments. RFLP data from 60 loci were collected from these cultivars. The average SMC among the southern elite lines was 0.72, the average SMC among the southern PIs was 0.76, and the average SMC between the members of the southern elite population and the PIs was 0.65. Diverse PIs with above average agronomic value were identified. These PIs may serve as sources of genetic diversity that can be exploited through simple breeding schemes.

Thompson et al. (1998) used 125 RAPD (Randomly Amplified Polymorphic DNA) primers to identify marker diversity in the 18 soybean ancestors which represents 86% of the northern soybean genetic base. Seventeen other PIs were selected for this marker analysis on the basis of their agronomic potential. The SMCs for all 35 genotypes were used to compute Euclidean distances. The Euclidean distances ranged from 0.26 for the Arksoy-Ralsoy pair of ancestors to 0.67 for CNS-PI384474 pair. For all possible pairings, the average pairwise genetic distance was 0.56. These SMCs indicated that the soybean DNA amplified by these RAPD primers was highly similar among the selected soybean lines.

George et al. (2003) analyzed the level of genetic diversity within and between Asian and North American soybean cultivars (NASC) by AFLP (Amplified Fragment Length Polymorphism) analysis. The genetic diversity and relationships were assessed among 35 North American soybean ancestors (NASA), 66 high yielding NASC, 59 modern Chinese cultivars, and 30 modern Japanese cultivars. Five AFLP primer-pairs produced 90 polymorphic (27%) and 242 monomorphic AFLP fragments. Average genetic distance (GD) among the 190 soybean genotypes was 8.1%, and the range of GD was 0.9 to 15.0%. There were no significant differences between the GD means of any of the four genotype groupings. The average GD was lowest among Japanese cultivars (6.3%) whereas the Chinese cultivars had the highest average GD estimate (7.5%). Genetic distance was 7.1% among the NASC, 7.3% among the NASA. The average GD between NASC and the Chinese cultivars was 8.5% and between NASC and the Japanese cultivars was 8.9%. The AFLP genetic distance clearly formed a distinct grouping of cultivars on the basis of origin. Intermating cultivars from these three major gene pools should provide new genetic recombinations to exploit in cultivar development programs.

Xu et al. (2003) evaluated the genetic diversity of 21 wild soybean accessions and 27 cultivated soybean landraces grown in China by using the RAPD method. The data showed that wild soybean has a higher genetic variation than cultivated soybean. All the accessions were
classified into two major clusters, corresponding to wild and cultivated varieties of soybean. Sub-clusters were identified within each species in agreement with their geographical origins demonstrating that an extensive geographical genetic differentiation exists in both wild and cultivated accessions.

O’Hara and Shimamoto (2002) studied the genetic diversity of wild soybean (*Glycine soja*). The genetic structures of 447 Japanese populations were analyzed. *Glycine soja* populations had a high degree of variation of isozyme loci as indicated in comparative studies (Hamrick & Godt, 1997). The GST (coefficient of gene differential) values among the sites within the district were particularly high, revealing the isozyme genotype was greatly different among site populations and homogeneous within the sites. O’Hara and Shimamoto (2002) concluded that the big genetic variation in *G. soja* should provide useful information for the improvement of the cultivated soybean.

Chen and Nelson (2004) used 92 *G. soja*, *G. max*, and semi-wild accessions from the USDA Soybean Germplasm Collection to quantify the genetic variation. Twenty phenotypic traits and 137 RAPD markers were evaluated. Three clusters defined by either phenotypic or RAPD data were highly consistent and strongly corresponded to *G. soja*, *G. max*, and semi-wild classifications. Based on the RAPD data, the semi-wild group has the smallest within-cluster genetic distance (0.107), and the *G. soja* group has the largest genetic distance (0.219). These results indicated that *G. soja* accessions have the greatest genetic diversity and semi-wild accessions have the least. The genetic distance between the semi-wild cluster and the *G. max* cluster (0.199) was the least distance among the clusters indicating *G. max* and semi-wild accessions are more closely related to each other than to *G. soja* accessions.

**QTL Mapping**

The availability of molecular markers has facilitated the identification, localization and genetic dissection of loci that control quantitatively inherited traits, such as yield (Tanksley, 1993). Molecular markers also offer a faster and more accurate approach to breeding, since selection can now be based on genotype rather than solely on phenotype. Finally, molecular markers provide a way to successfully tap the genetic diversity available in exotic germplasm for crop improvement. Molecular markers now allow the isolation of beneficial genes that are often tied up in unfavorable linkages in exotic germplasm, and their transfer into elite commercial germplasm (Tanksley and McCouch, 1997).
Two groups (Kearsey and Hyne, 1994; Hyne and Kearsey, 1995; Wu and Li, 1994, 1996) proposed a very simple and powerful regression method that simultaneously considers all of the markers on a single chromosome. These authors refer to this method as Marker Regression or single-marker regression. Single-marker regression is based on comparisons between marker genotypic means through a simple regression for trait on coded marker genotype. The single-marker regression is done by analyzing one marker at a time. The QTL is determined to be located near a marker if phenotypic values for the trait are significantly different among the marker genotypes. Single-marker analysis is simple. Gene orders and a complete linkage map are not required. The disadvantages of the single-marker regression are (Liu, 1998): 1) the putative QTL genotypic means and QTL positions are confounded. This confounding causes the estimator of QTL effects to be biased and the statistical power to be low, particularly when linkage map density is low. 2) QTL position cannot be precisely determined, due to the nonindependence among the hypothesis tests for linked markers that confound QTL effect and position.

Lander and Botstein (1986b, 1989) proposed an interval mapping approach to locate QTL. The interval mapping approach is based on the joint frequencies of a pair of adjacent markers and a putative QTL flanked by the two markers. Simple interval mapping of QTL needs a complete genetic map and a QTL effect can be unconfounded from its distance to a marker, since both are independently estimated in a marker-flanked genomic segment. Problems exist for the simple interval mapping approach (Benhui Liu, statistical genomics, 1998): 1) The number of QTL cannot be resolved, 2) The locations of QTL are sometimes not well resolved and the exact positions of the QTL cannot be determined, 3) The statistical power is still relatively low. The assumption for simple interval mapping is that a single segregating QTL influences the trait. The most important reason for these problems is that interval mapping is not independent for different segments if more than one QTL exists.

Composite interval mapping (CIM) is a combination of simple interval mapping and multiple linear regressions (Zeng, 1993, 1994). It adds background loci to simple interval mapping (Jansen, 1993; Zeng, 1993) to control for trait variance arising from those markers, while testing other marker-flanked segments for a possible QTL-level effect. The CIM has advantages over simple interval mapping and single marker analysis (Zeng, 1994). For CIM, the
markers can be used as boundary conditions to narrow down the most likely QTL position. The resolution of QTL locations can be greatly increased. The background markers used in this study were significant markers identified with the Map Manager marker regression analysis. The background markers can be in different or in the same linkage groups. To control for unlinked QTL, select one marker from each chromosome (other than the chromosome being mapped) which has loci significantly associated with the trait. To control for linked QTL, include one marker between the QTL being mapped and the region of the chromosome suspected of containing another QTL (Zeng, 1993 and 1994).

**Advanced Backcross**

The most commonly used QTL mapping population are F$_2$, BC$_1$, recombinant inbred lines and doubled haploid lines. These populations have been used because: 1) they allow genetic mapping to occur simultaneously with QTL analysis; 2) there is nearly equal contribution of alleles from each parent, maximizing the analysis of the genetic differences between the two parents. But these types of populations are not ideally suited for introgression studies from the wild species, because: 1) QTL analysis in these early generation populations tends to separate QTL discovery from the incorporation of QTL into advanced breeding lines; 2) It is difficult to map QTL from early generation of crosses between soybean and *G. soja*, because of the poor general agronomic performance of lines, such as seed shattering.

There are several population development methods which can solve these problems. Wehrhahn and Allard (1965) were probably the first plant geneticists to recognize the value of the inbred backcross method. In this method BC$_2$ or BC$_3$ individuals are inbred by single seed descent resulting in homozygous lines with isolated introgressed segments. Inbred backcross populations have been used to introgress unadapted germplasm (Sullivan and Bliss, 1983; Hartman and St. Clair, 1998) and as population for QTL identification (Butruille et., 1999).

Tanksley and Nelson (1996) suggested a strategy termed advanced backcross (AB) QTL analysis. The AB-QTL method is really the inbred backcross method of Wehrhahn and Allard (1965), but includes molecular marker screening of the backcross progeny to identify the specific donor-parent-derived genomic segments influencing yield and other traits. AB-QTL analysis has been used for QTL identification in the population developed from the cross between wild and cultivated varieties.
Doganlar et al. (2002) compared QTL identification in inbred backcross and advanced backcross populations from the same introgression program. In the advanced backcross BC\textsubscript{1} and BC\textsubscript{2} populations 17 and 33 QTL were identified for five traits, respectively. Only 11 QTL were detected in the inbred backcross population, with only five of those detected in at least one of the advanced backcross populations. The percentage of the phenotypic variation explained by each QTL was greater in advanced backcross populations than in inbred backcross populations, and a greater number of QTL with favorable effects from wild parent alleles were identified in advanced backcross compared to inbred backcross populations.

This method was first explored in tomato (*Lycopersicon esculentum*). The concept of mining genes in wild relatives has been demonstrated successfully many times (Grandillo and Tanksley, 1996; Tanksley and Grandillo, 1996; Fulton et al., 1997; Bernacchi and Beck-Bunn, 1998b; Monforte and Tanksley, 2000). The research groups of Tanksley (1996) and Bernacchi (1998a,b) have conducted advanced backcross QTL analysis of crosses between cultivated *Lycopersicon esculentum* and the wild species *L. hirsutum* and *L. pimpinellifolium*. In the cross with *L. hirsutum* they identified 12 QTL for yield. Eleven of these identified the *L. hirsutum* allele as negative. But QTL ydt4.1 on chromosome 4 was different. Lines carrying the *L. hirsutum* allele yielded 15% more on average than lines homozygous for the *L. esculentum* allele (Bernacchi and Beck-Bunn, 1998b). In the *L. pimpinellifolium* population one out of six QTL for yield showed positive yield benefits associated with the wild species allele (Tanksley and Grandillo, 1996). In both populations a yield QTL was identified at ydt3.1 on chromosome 3. With *L. pimpinellifolium* the wild allele was positive for yield while in the *L. hirsutum* introgression population the allele from *L. hirsutum* was negative for yield. Thus it should be possible to incorporate a favorable allele for yield from each of these germplasm accessions into cultivated tomato.

Rice (*Oryza sativa*) is another crop which has recently been accessed for favorable yield alleles from a wild species. Xiao et al (1996) crossed and backcrossed a wild accession of rice (*Oryza rufipogon*) to produce 300 BC\textsubscript{2} families. Four percent of the testcross families out yielded the control hybrid by at least 30%. QTL on chromosome 1 and chromosome 2 increased yields by 18 and 17% over the elite hybrid. Xiao et al. (1998) found 51% beneficial QTL underlying important agronomic traits originating from *Oryza rufipogon*.  

16
Moncada et al. (2001) used the same *O. rufipogon* accession to create a population with the upland rice variety Caiapo as the cultivated parent. From 40 BC1 plants, 274 BC2F2 families were screened with 125 markers at approximately 10 cM intervals and tested in two environments. Two yield and 13 yield component QTL were identified. Of the *O. rufipogon* alleles, 56% were associated with the favorable agronomic response. The two *O. rufipogon* alleles at the yield QTL increased yield. For all seed size QTL the *O. rufipogon* alleles produced heavier seeds despite the *O. rufipogon* accession having 15% lighter seeds. Even though *O. rufipogon* is phenotypically inferior for 7 of the 8 traits studied, transgressive segregates superior to Caiapo were identified.

Thompson et al. (2003) created a BC2F2 population using the same *O. rufipogon* accession and the US variety Jefferson. Out of this population, 258 BC2F2 families from 33 BC1 plants were screened with 153 markers at approximately 10 cM intervals and tested in three environments. Seventy-six QTL were identified. Of these *O. rufipogon* alleles, 53% were associated with the favorable agronomic response. For five yield and 39 yield component QTL the *O. rufipogon* alleles were positive at four and 17 QTL, respectively. These results show that favorable *O. rufipogon* QTL alleles were expressed in different genetic backgrounds and environments.

Brondani et al. (2002) identified favorable wild species QTL alleles for tiller number and panicle number in a 96 family BC2F2 population with *O. glumaepatula* as the wild species donor parent. Overall 16% and 9% of the *O. glumaepatula* alleles were favorable for the two environments tested.

Brinda and Brown-Guedira (2003) developed 189 BC2F2:4 lines using a synthetic wheat *Ae. tauschii* introgression line and the variety Karl 92 and conducted tests in four environments. The favorable allele came from the synthetic wheat introgression line for 24 QTL with the most relevant alleles increasing yield 7.7%.

**Soybean Molecular Marker Map and Yield QTL**

Genetic linkage maps serve the plant geneticist in a number of ways, from marker assisted selection in plant improvement to map-based cloning in molecular genetic research. The first two molecular maps of soybean (RFLP-restriction fragment length polymorphism) were constructed using *G. soja × G. max* interspecific crosses in order to produce enough polymorphisms to construct the map (Shoemaker and Olson 1993; Rafalski and Tingey 1993).
Soybean was not highly polymorphic for this marker class although over 500 RFLP markers were eventually mapped (Cregan et al., 1999). Furthermore, the RFLP assay is time consuming and based on the maintenance and exchange of specific DNA probes. Another factor that complicates the use of RFLP markers in soybean is the detection of multiple DNA fragments (i.e. multiple loci) with most probes. One fragment in a multiple banding pattern may segregate in one population and a different or an additional fragment in another. The multiplicity of RFLP loci can make RFLP linkage maps ambiguous with respect to RFLP locus identity, and often precludes the use of such loci for the evaluation of linkage group homology among different maps.

One possible solution to this complexity is the use of single locus DNA markers with multiple alleles. In soybean, the highly polymorphic nature (i.e. multi-allelism) of simple sequence repeat (SSR) DNA markers is well known. Multiple allele molecular markers are much more useful than dimorphic markers when tracking the fate of genomic segments in multi-parent pedigrees and in multi-parent mated populations. Another virtue of SSR markers is their simplicity. It requires less genomic DNA and assays are less time consuming.

The soybean SSR map is now well developed. A total of 606 SSR loci were mapped in one or more of three populations (Cregan et al., 1999): the USDA/Iowa State G. max × G. soja F2 population, the Univ. of Nebraska Clark × Harosoy F2 population, and the Univ. of Utah Minsoy × Noir1 recombinant inbred population. Cregan et al. (1999) aligned the 20 linkage groups (LG) derived from each of the three populations into a consensus set of 20 homologous linkage groups presumed to correspond to the 20 pairs of soybean chromosomes.

More recently, Song et al. (2004) mapped 420 newly developed SSRs in one or more of the five soybean mapping populations: Minsoy × Noir 1, Minsoy × Archer, Archer × Noir 1, Clark × Harosoy, and A81-356022 × PI468916. They used the JoinMap software package to combine the five maps into an integrated genetic map spanning 2523.6 cM of Kosambi map distance across 20 linkage groups that contain 1849 markers, including 1015 SSRs. The number of SSRs added to each linkage group ranged from 12 to 29.

The availability of a molecular genetic map saturated with highly informative, PCR-based, single-locus, multiple-allele SSR markers is a great utility for this research. In soybean, many of the beneficial genes that have been extracted from exotic or unadapted germplasm using molecular markers are involved in pest and disease resistance. Examples of
QTL for pest resistance from exotic germplasm include genes for resistance to soybean cyst nematode (Concibido and Denny, 1994; Webb and Baltazar, 1995; Concibido and Denny, 1997), peanut root-knot nematode (Tamulonis and Luzzi, 1997), and corn earworm (Rector and All, 2000; Terry and Chase, 2000). For disease resistance, QTL from exotic soybean relatives have been reported for brown stem rot (Klos and Paz, 2000), soybean mosaic virus (Hayes and Ma 2000) and phytophthora root rot (Hegstad and Nickell 1998). Other studies resulted in the discovery of QTL for oil, protein and other seed quality traits (Diers and Keim 1992; Brummer et al.1997; Sebolt et al. 2000).

Research for yield QTL originating from exotic germplasm is not a new concept in soybean (Lark and Chase, 1995; Orf and Chase, 1999a, b). However, none of these studies addressed the issue of adaptability of exotic yield QTL across genetic backgrounds.

Orf and Chase (1999a,b) used lines from Minsoy × Noir 1 (MN), Minsoy × Archer (MA), and Noir 1 × Archer (NA) populations and identified five yield QTL with RFLP and SSR markers. No major yield QTL was found in the MA population. Two yield QTL were found in the MN population. Three yield QTL were found in the NA population. The yield locus in LG U11, segregating in the MN population, was associated with a major maturity QTL.

Specht et al. (2001) used the genotypic data of Orf and Chase (1999) from the Minsoy × Noir 1 population and identified six yield QTL under water stress conditions. The strongest QTL for yield, based on the variance explained and the effect of the Minsoy allele, was located near the top of linkage group U11-M. This QTL had been shown to be a major QTL for maturity and yield. In each year of the study, the U11-M QTL accounted for about 23 to 29% of the phenotypic variation in maturity and about 33 to 38% of the variation in seed yield. The second strongest QTL with an impact on yield was also a maturity locus located near the center of LG U09-C2. This QTL accounted for 13 to 15% of phenotypic variance in maturity and about 7 to 13% of the variation in yield.

Yuan et al. (2002) used SSRs in the Essex × Forrest and Flyer × Hartwig populations and identified five QTL for yield. High yielding and SCN resistant transgressive segregants were not observed. In E × F population, three QTL for yield were found. They were identified by Satt294 on linkage group C1, Satt440 on LG I, and Satt337 on LG K. In the F × H population, two QTL were identified. They were identified by Satt326 and Satt539 on LG K. Both populations
revealed a yield QTL in the interval (5cM) between Satt337 and Satt326. These populations may share a common allele for yield in this region.

Concibido et al. (2003) used SSR and AFLP markers and the advanced backcross method of QTL mapping described by Tanksley and Nelson (1996) and identified a yield QTL in a HS-1 × PI 407305 population. In their study, 265 BC2F2 derived lines were tested in ten environments and screened with 212 AFLP markers. The yield QTL named as AFLPU3944 on linkage group B2 provided a significant 9.3% yield increase across the 10 environments. This yield QTL provided a significant yield increase in 6 of the 10 environments. They concluded that with the availability of molecular markers and the existence of programs like SAGE (Soybean Asian Germplasm Evaluation), the use of exotic germplasm is more attractive than previously. Molecular markers allow the delivery of beneficial genes from the wild into commercial varieties with a tolerable amount of linkage drag.

Smalley et al. (2004) conducted experiments to prove that PIs possessed favorable alleles for yield. Allele frequencies were measured with SSR markers in three populations: AP10, AP12, and AP14 that differed in their percentage of PI parentage. AP10 had 40 PI parents, AP12 had 40 PIs and 40 elite parents, and AP14 had 40 elite parents. Four cycles of recurrent selection for yield had been conducted in the three populations. Allele frequencies of the highest-yielding C4 lines in the three populations were compared with the parents used to form the populations of the initial cycles. Fivty-four SSRs were associated with 43 yield QTL. There were 16 alleles at 15 SSRs unique to the PIs, 9 alleles at 9 SSRs unique to the elites. These results indicated that some of those unique PI alleles remained after the fourth cycle of selection and increased appreciably in frequency in the highest-yielding C4 individuals.

Kabelka et al (2004) evaluated two soybean PIs as sources of alleles for the enhancement of seed yield in North American cultivars. They developed a population by using the F5-derived lines from the cross of BSR101 x LG82-8379. Ancestry of BSR101 includes nine of 10 major ancestral lines that contributed to the commercial gene pool of North America. LG82-8379 was selected from a cross between PI68508 and FC04007B. Fifteen QTL were found associated with seed yield. Nine of the yield QTL were LG82-8379 alleles and associated with yield increases of 1.7 to 5.4%. Six of the BSR101 alleles increased yield by 2.4 to 4.4%. Among these QTL, Satt358 on linkage group O was significant in the combined analysis across sets of this population. For this QTL, FC04007B allele enhanced yield 47 kg ha⁻¹ and increased seed protein.
concentration 3 g kg$^{-1}$. To date, no other studies have detected a yield QTL at this chromosomal location. From this study, they concluded that soybean PIs contain QTL alleles that may be of benefit to North American soybean cultivars.

Wang et al. (2004) attempted to map QTL from *G. soja* that could improve the crop. Crosses were made between IA2008 (*G. max*) and PI468916 (*G. soja*). Five populations of BC$_2$F$_4$-derived lines were developed. They found four yield QTL across environments on linkage groups C2, E, K, and M. But for these QTL, the *G. max* marker allele was associated with significantly greater yield than the marker allele from *G. soja*. They presented several possible reasons for these results. First, *G. soja* parent PI468916 lacks positive yield QTL that could improve IA2008. Another possibility is that the *G. soja* parent did have one or more positive QTL, but they were not detected in this study. This may have occurred if there were no segregating markers near the QTL or if polymorphic markers were present, but the populations used in the study were of insufficient size to detect the QTL.

**Why Look For Seed Filling Period QTL**

Crop yield can be divided into its components, seeds per unit area and seed size. Seed number is the yield component that accounts for most of the environmental variation in yield of grain crops. Seed size is not as closely associated with environmental variation in yield as is seed number. The relationship of these components to yield is confusing and frequently contradictory. In order to reduce this confusion, we can consider the fundamental processes underlying the yield components, i.e. seed growth rate (SGR) and seed filling period (SFP) (Egli, 1998).

Genetic differences of SGR are regulated by the seed through the number of cells in the cotyledons or endosperm, not by the ability of the mother plant to supply assimilate to the seed. Seed growth rate can be regulated by the seed through its ability to supply assimilate to the seed. Genetic differences in SGR are yield neutral (Egli, 1998).

Genotypic differences in SFP have been found in many crops suggesting that SFP is under genetic control (Egli, 2004). Experimental data show that this is true for soybean SFP (Gay et al., 1980; Zeiher et al., 1982; Boon-Long et al., 1983). The positive relationship between genetic variation in SFP and yield is predicted by theoretical analysis and is supported by experimental data. Longer seed fill durations in crop simulation models were associated with higher soybean yield (Boote et al., 2001). Plant breeders inadvertently lengthened the seed filling period when selecting for higher yield (Egli, 2004). Positive association was reported among
soybean cultivars by Dunphy et al. (1979) and in comparisons of old and new soybean cultivars (Gay et al., 1980; McBlain and Hume, 1980). Smith and Nelson (1986) reported that modification of SFP by direct selection has also increased soybean yield. Plant breeders modified seed-filling duration by direct-selection in soybean (Smith and Nelson, 1987; Pfeiffer and Egli, 1988). Higher yields must come from either more seeds per unit area or from a longer seed filling period (Egli, 1998). Lengthening the seed filling period may hold more potential for future yield increase.
Figure 1. Graphic representation of soybean yield (kg/ha) in the United States, Brazil, China, Argentina, and India in 1961-2004 (source: www.fao.org)
CHAPTER THREE

THESIS RESEARCH OBJECTIVES

A mating of the *Glycine soja* germplasm accession PI 245331 (donor parent) with the high yielding *Glycine max* cultivar 7499 (recurrent parent) was used to produce a 147-member BC$_2$F$_4$ population.

Our research objectives were to:

(1) Phenotype in the 2003 and 2004 growing seasons the BC$_2$F$_4$ BILs (backcross inbred lines) in two different locations for yield, seed filling period, maturity, height, lodging and seed size; genotype these BILs with about 120 SSR markers spaced about 25cM apart (on average) in a 3000-cM soybean genome.

(2) Map yield, seed filling period, maturity, height, lodging, and seed size QTL in a 147 line BC$_2$F$_4$ *G. max × G. soja* inbred backcross introgression population. Determine if an advanced backcross QTL analysis of this mating could detect favorable *G.soja* alleles at QTL governing yield, and other traits.

(3) Verify the QTL in a second similarly derived population.
CHAPTER FOUR
MATERIALS AND METHODS

Parental Germplasm and Population Development

The *G. soja × G. max* inbred backcross population was developed using the parents *G. soja* PI 245331 and the *G. max* variety 7499. PI 245331 is a maturity group X accession from Taiwan which was identified by Maughan et al. (1996) as being more diverse from *G. max* than the remaining *G. soja* group. They reported that this accession was previously classified as *G. formosana* and may represent a distinct sub-species of soybean.

Variety 7499 soybean was developed by the Kentucky Agricultural Experiment Station (Pfeiffer, 2000). It was released in 1998 because of its superior seed yield in both full-season and double-crop plantings compared with the public cultivars of similar maturity most widely grown in Kentucky. Variety 7499 is an indeterminate, maturity group IV cultivar. Using photoperiod (12 hr day; 12 hr night) manipulation to achieve flowering, PI 245331 was crossed in the greenhouse in Feb. 1998 with variety 7499 (Table 1). After selfing the F1 in the field summer 1998, 87 F2 progeny with an acceptable flowering response (greenhouse 14.5 hr day; 9.5 hr night) were backcrossed to 7499 (March, 1999). The BC2 population was derived from crossing (July 1999) onto 89 BC1 plants tracing back to 53 F2 plants. These BC2 progeny were advanced by single seed descent to the BC2F4 generation to produce the two inbred backcross populations (BIL).

BC2F1 plants were grown in the greenhouse in winter 1999-2000. BC2F2 plants were grown as BC2F1:2 families in summer 2000. Two plants were advanced by single seed descent (Brim, 1966) from each family to create two related populations A and B. Single seed descent was continued in both populations in the greenhouse in winter 2000-2001. Individual BC2F4 plants were harvested in October 2001.

In 2002, the lines were grown in a five replication hill plot test that also served as the seed increase generation. Phenotypic traits, such as height, maturity, lodging and yield were measured. Based on the 2002 maturity data, lines were separated into three maturity sets in both the A and B tests in 2003 and 2004. Population A has 147 lines and was used to build the QTL map, population B has 148 lines and was used to verify the results from A.
Phenotypic Trait Measurement

Performance trials of the 295 BILs were conducted in 2003 and 2004 at Lexington and Princeton, KY. The soil types were Maury silt loam (Typic Paleudalfs) at Lexington, and Crider silt loam (Typic Paleudalfs) at Princeton.

Both test A and test B lines were planted according to a split plot design repeated over locations with a one-way treatment classification (genotype), three maturity sets per environment and two replications per maturity set at each of the two locations in both years. Plot size was two 6m rows with 0.76m between rows for each line. The plot was end trimmed to 5m prior to harvest.

Six soybean traits were measured: yield, seed size, height, lodging, seed filling period and maturity. Yield (kg/ha) was measured within each line. Height (cm) was measured as the distance from the soil to the tip of the main stem at maturity. Duration of seed filling was measured at Lexington as the number of days from R5 (beginning seed fill) date to the date of R7 (physiological maturity) (Fehr and Caviness, 1977). The growth stages were measured on 10 consecutive plants in one row of the two row plots and the plot was considered to be at a given growth stage when five or more of the 10 plants reached that stage. The 10 plants were identified prior to growth stage R5 and all measurements were made on the same 10 plants. The plots were observed once a week before R5 and three times a week between R6 to R7. (Egli, et al., 1984). The R5 date was determined by interpolation of the weekly growth stage data collected prior to R5. Prior to harvest, the plots were periodically evaluated to establish a mean harvest maturity date (d from 30 August to the date when 95% of the pods had reached a final pod color). Lodging is based on a visual rating between one to five as follows: 1, almost all plants erect; 2, either all plants leaning slightly or a few plants down; 3, either all plants leaning moderately or 25 to 50% of the plants down; 4, either all plants leaning considerably or 50 to 80% of the plants down; and 5, all plants prostrate. (Wilcox, 2001). Just prior to harvest, pod shattering in a plot was visually scored (scale of 0 to 10, where 1 indicated up to 10% shattered, 2 up to 20%, etc.). The plots were harvested with a plot combine. The harvested seed was weighed and its seed moisture measured to estimate plot seed yield (kg/ha adjusted to the standard 13% seed water content). A 100-seed sample was drawn and used to obtain a 100-seed weight (g adjusted to a 13% seed water content).
Statistical Analysis

Analysis of variance (ANOVA) was conducted on each measured trait, using the PROC MIXED model of version 8 of PC SAS for Windows (SAS, 1999). Locations and years were treated as fixed effects; replication, and entries (147 BILs) as random effects.

Agronomic traits of the A-population were analyzed for both single environment and combined environments. For the combined environments analysis, maturity sets, environments, and environments × sets were considered as fixed effects. Replications within environment, lines within maturity sets, and environments × lines within sets were considered as random effects. Each location-year combination was considered an environment in the analysis. Significance was determined by an F test as described by McIntosh (1983), with the numerator and denominator degrees of freedom approximated as described by Satterthwaite (1946). For single environments, effects of maturity sets were considered as fixed. Effects of replications, lines within maturity sets, and replications × sets were considered as random.

Line means across environments were computed for each trait. Using PROC CORR in SAS, Pearson correlation coefficients among all traits were calculated both within and pooled over maturity sets using the line means across environments.

In 2003 at Lexington, the plots of the QTL mapping (A) population were infected with tobacco ringspot virus producing bud blight infection with an accompanying pod abortion. Extensive yield loss was observed on many plots, and a total yield loss was observed on 20 of the 147 plots. In order to test the reliability of the Lexington 2003 data, a correlation analysis was conducted for the line means between environment 03Lex and the other three environments: 04Lex, 03Prin and 04Prin.

Since seed filling period and maturity data notes were taken only in 03Lex and 04Lex, to improve the effectiveness of the data, plots with severe bud blight disease were removed. Disease was serious on border plants in a row. Healthy plants were chosen from the middle of the row and seed filling period and maturity notes were taken only on the healthy plants. In order to test the reliability of the 03Lex data, a correlation analysis was conducted on the line means between environment 03Lex and 04Lex.

Broad-sense heritability of each trait was calculated by the variance component method on a line-mean basis according to Fehr (1987).
$H^2 = \frac{\sigma_g^2}{(\sigma_e^2/rt + \sigma_{ge}^2/t + \sigma_g^2)}$

$\sigma_g^2$: genetic variance; $\sigma_e^2$: experimental error; $\sigma_{ge}^2$: genotype by environment interaction

$r$: number of replications; $t$: number of test environments

**Leaf Collection and DNA Isolation Procedures**

Leaves were collected from plants of the 295 lines (including population A and B) grown in the greenhouse in 2002. Leaf material was also collected from the parents PI 245331 and 7499. The leaf samples were stored at -20 °C prior to and after lyophilization. Leaves were freeze dried and powdered. DNA from the 295 lines was extracted using the Qiagen DNAeasy mini-prep kit from Qiaexinc. The DNA was quantified using a fluorometer and diluted to 10 ng/ul. 10 ul of this DNA suspension was run on a 1% agarose gel to check for DNA quality and to verify the concentration.

**Phenotypic Marker**

Several vegetative and reproductive pigmentation traits were segregating in the BC$_2$F$_4$ population. PI 245331 has black pods. This trait is controlled by a single genetic locus that is expressed as phenotypes black pod color (L1L1) vs. brown pod (l1l1). This phenotypic marker was mapped using its gene symbol L1.

**SSR Markers**

The two parents were screened for polymorphisms at 534 SSR marker loci. Out of these SSR markers, 284 polymorphic markers were identified. Selecting markers with reasonable spacing resulted in 130 markers, placed at an average interval of 25 cM, to be used in the initial QTL discovery analysis. Of those 130 markers, 5 produced a multiple banding pattern that could not be reliably scored in the BIL screen, and 5 could not be scored due to weakly expressed banding patterns, or a lack of amplification. In the end, 147 BILs were genotyped for 121 markers including 120 SSRs and one phenotypic marker (L1). A minimum of four markers was evaluated on each linkage group; the mean number of markers per linkage group was 6. These 121 markers covered a total length of 2215 cM or 74% of the soybean genome (estimated to be approximately 3000 cM), with an average interval size of 18.3 cM. I have not identified SSR markers to place in the 64 cM interval on chromosome A2 or to more adequately cover the 61 cM gap at the bottom of chromosome N.

31
PCR and Gel Electrophoresis

PCR mixes consisted of between 15.0 and 30.0 ng template DNA, 1.5 mM MgCl₂, 0.15 μM forward and reverse primers, 200 μM of each nucleotide, and 1.0 unit taq polymerase. The PCR thermocycling profile was 2 minutes at 94°C, then 34 cycles of: 25 seconds at 94°C, 25 seconds at 68°C, and 25 seconds at 47°C. Amplified PCR fragments were separated on either metaphor agarose or polyacrylimide electrophoresis depending on the size of the polymorphism between the parents. Polymorphisms greater than 10 base pairs were run on 3% metaphor agarose stained with ethidium bromide gels for 4 hours at constant 70 volts. Polymorphisms that were less than 10 base pairs were run on 6% non-denaturing polyacrylimide gels at 300 volts for 4 hours. Polyacrylamide gels were stained with ethidium bromide after electrophoresis (Stefaniak et al., 2006).

Marker Scoring and Segregation Analysis

At each SSR locus, the homozygous variety 7499 marker allele was assigned the genotypic code A, the homozygous PI 245.331 marker allele was assigned the genotypic code B, and the code H was assigned to the heterozygote. If a BIL phenotype was not discernible or was missing, it was assigned a dash (-) to indicate a missing data point. Of the 147 population A BILs × 121 loci=17787 potential data points, only 245 were coded as missing.

The segregation ratio for the two parental marker alleles at each marker locus was expected to be 86% M: 3% H: 11% S, or if heterozygotes are ignored, a ratio of 7.86 M: 1 S was expected (M= G. max, S= G. soja). At each of the 121 markers, goodness of fit of the expected genotypic ratio (7.86M: 1S) to the observed marker segregation data was evaluated by the Chi-square test (α=0.05).

Linkage Map Construction

Map Manager QTXb.20 software (Meer et al., 2003, Manly et al., 2001) was used to create a genetic linkage map for the BC₂F₄ population. Map Manager can construct a linkage map from an unselfed BC₃ population (7A:1H segregation ratio), and thus can also construct a linkage map from a selfed BC₂ population (7A:1B ratio, with a few residual H) by a simple recoding of the H genotypes to a dash (-, missing data) and the B genotypes to an H (Chaky, 2003). The raw marker data file was formatted in the standard manner (marker genotype codes of A or H at the intersection of marker rows and BIL columns). The formatted raw data file was imported into Map Manager (in TXT file with TAB delimited format). For linkage map
construction, the Map Manager parameters and settings were: 1) Linkage Evaluation set to advanced backcross 3, 2) Search Linkage Criterion set to P=0.001, 3) Map Function set to Haldane, and 4) Cross Type set to Line Cross. The Make Linkage Groups command resulted in the 121 markers coalescing into 20 different linkage groups (LGs). The marker groups were consistent with the expected grouping based on the LGs of SoyBase (2003). The markers in each linkage group were then subjected to the Ripple command to establish the most probable final marker order. Finally, the 20 LGs were aligned and named, on the basis of SSR marker homology, to the LGs defined by Cregan et al. in SoyBase (2003). Those 20 LGs are displayed in Figure 2.

QTL Analysis

QTL analysis was performed using the marker regression (MR), simple interval mapping (SIM), and composite interval mapping (CIM) functions of Map Manager QTxb.20 software (Meer et al., 2003). The input file used in Map Manager of BIL phenotype data for all the phenotypic traits was created in the trait-row x BIL-column format (in TXT file with TAB delimited format) (Chaky, 2003). This trait file was imported into Map Manager using the Import Trait Text option.

The QTL analysis results reported include: 1) the genomic map position of each LRS score peak (i.e., the QTL), 2) the LOD score of putative QTL, 3) the sign and magnitude of the trait additive genetic effect if the B allele of the donor parent *G. soja* were to be substituted into a BIL homozygous of the A allele of recurrent parent *G. Max* (note: The Map Manager additive value was halved to convert the BC$_3$ estimate to a BC$_2$ self estimate), and 4) percentage of the total trait variance explained by a single QTL model.

The Map Manager MR procedure to identify significant markers on a single locus basis was performed as follows: 1) Select Marker Regression (in QT menu), 2) Select trait (e.g., yield), 3) Select chromosomes to be included in the regression, and 4) Select “none” for trait variances option. The output subsequently lists marker loci that are significant. Map Manager MR analysis also has the option to include significant markers as background markers, but this option was set to none in the initial MR and SIM analysis.

The SIM procedure was executed as follows: 1) Select Interval Mapping (in QT menu), 2) Select Map locus for trait (e.g., yield), 3) Select interval (e.g., 1cM steps), and 4) Select chromosome to map. The CIM procedure is similar to SIM, except that significant markers are
typically chosen to serve as background markers. In many instances, background markers will simply be the more significant of the two markers flanking the MR or SIM detected QTL.

The output from Map Manager includes a report, listing the names of the marker loci, map position, LRS at each position, percent trait variance explained, sign and magnitude of the additive effect, as well as a graphical representation of the mapped QTL. A LOD score maximum, if statistically significant, was assumed to be the mean position of a QTL. A one-LOD fall-off from each side of that peak LOD was used to estimate the left and right confidence intervals (in cM) of the presumed QTL at that peak.

A genomewide likelihood ratio statistic (LRS) criterion for evaluating the statistical significance of QTL effects for each trait was estimated for MR, SIM, and CIM by permutation (Churchill and Doerge, 1994). For the BC2:F4 population, this LRS criterion was determined to be about 13.8 (i.e., LOD score=LRS(13.8) * 0.217=3.0) for the MR analyses, based on N=1000 permutation tests conducted with the yield, height, lodging, seed weight, seed filling period and maturity data. QTL with the LOD>2.0 but less than 3.0 were used as background markers in the CIM analysis.

For SIM and CIM QTL analysis with combined data, the 1000 permutation tests were used to define an empirical detection threshold at an experiment-wise P-value<0.01 (corresponding to a SIM mapping LOD>3.6 or a CIM LOD>3.8) and P<0.05(corresponding to a SIM and CIM LOD>3.0). For single environment SIM analysis, P<0.05 was used.

**QTL Confirmation**

The major objective in this study was to identify trait-enhancing alleles from the wild soybean relative PI 245331 for yield and yield components. To confirm in a second test the identified QTL associated with PI 245331 alleles found in population A, another 148 BC2:F4 lines (population B) were used. Population B and population A are related with one line in each population tracing back to the same BC2 plant.

**Marker Analysis in B-population**

The 148 lines in population B and the two parents were screened with markers Satt511, Satt529 and Satt335. The PCR mixes and PCR thermocycling profile were the same as those used for population A. Amplified PCR fragments were separated on metaphor agarose for markers Satt529 and Satt335 and polyacrylimide electrophoresis for marker Satt511. Marker scoring was the same as that in population A.
Statistical Analysis in B-population

ANOVA were conducted in the B-population on both single environments and data combined from the four environments to calculate the QTL effects and QTL × Environment interaction effects, following the procedures of Sari-Gorla et al. (1997).

In 2003 at Lexington, the plots of population B were not infected with tobacco ringspot and soybean bud blight. All four environments were used for the confirmation analysis.

Seed yield and seed filling period traits and their marker data in the B-population were analyzed with ANOVA by the PROC MIXED functions of SAS (Statistical Analysis System version 8.0, SAS Institute, Cary, NC) for both single environment and combined environments.

Because two QTL were identified for seed yield, QTL1 (Satt511), QTL2 (Satt529), QTL1×QTL2, and their environmental interaction effects were calculated. For the combined environments analysis, maturity sets, environments, environment × set, QTL1, QTL2, QTL1×QTL2, QTL × Environment, QTL × Set and QTL × Environment × Set were considered as fixed effects. Replications within environment, replication × set, lines within QTL and maturity sets, and environment × line × set were considered as random effects. Each location-year combination was considered an environment in the analysis. Significance was determined by an F test. LSmeans (least squares means) were used to test the allelic means and the four class differences (mm, ms, sm and ss) for combined allelic type across the testing environments. For the single environment analysis, maturity sets, QTL1, QTL2, QTL1×QTL2, and QTL × set were considered as fixed effects. Replications, replications × sets, and lines within QTL and maturity sets were considered as random effects.

For seed filling period, only one positive QTL was identified. QTL and QTL by environment interaction effects were calculated. For the combined environments analysis, maturity sets, environments, environments × sets, QTL, QTL × environments, QTL × sets and QTL × environments × sets were considered as fixed effects. Replications within environments, replications × sets, lines within QTL and maturity sets, and environments × lines × sets were considered as random effects. Each location-year combination was considered an environment in the analysis. Significance was determined by an F test. For the single environment analysis, maturity sets, QTL, QTL × sets were considered as fixed effect. Replications, replications × sets, lines within QTL and maturity sets were considered as random effects.
Table 1. Development of 294 BC₂F₄ lines from a 7499 (G. max)×PI 245331 (G. soja) mating

<table>
<thead>
<tr>
<th>Generation</th>
<th>Crosses or Selfs</th>
<th>Season/Year</th>
</tr>
</thead>
<tbody>
<tr>
<td>Plant grown Seed harvested</td>
<td>7499(max)×PI245331 (soja) F₁</td>
<td>Winter GH 1997-98</td>
</tr>
<tr>
<td>Plant grown Seed harvested</td>
<td>F₁ 87F₂</td>
<td>Summer 1998</td>
</tr>
<tr>
<td>Plant grown Seed harvested</td>
<td>7499 × 87F₂ 89BC₁</td>
<td>Winter GH 1999</td>
</tr>
<tr>
<td>Plant grown Seed harvested</td>
<td>7499 × 89BC₁ BC₂</td>
<td>Summer 1999</td>
</tr>
<tr>
<td>Plant grown Seed harvested</td>
<td>BC₂F₁ BC₂F₁:2</td>
<td>Winter GH 2000</td>
</tr>
<tr>
<td>Plant grown Seed harvested</td>
<td>BC₂F₂ families BC₂F₂:3</td>
<td>Summer 2000</td>
</tr>
<tr>
<td>Plant grown Seed harvested</td>
<td>BC₂F₃ BC₂F₃:4</td>
<td>Winter 2000-01</td>
</tr>
<tr>
<td>Plant grown Seed harvested</td>
<td>BC₂F₄ lines BC₂F₄:5 progenies</td>
<td>Summer 2002</td>
</tr>
<tr>
<td>Plant grown Seed harvested</td>
<td>BC₂F₄:5 lines BC₂F₄:6 progenies</td>
<td>Summer 2003</td>
</tr>
<tr>
<td>Plant grown Seed harvested</td>
<td>BC₂F₄:6 lines BC₂F₄:7 progenies</td>
<td>Summer 2004</td>
</tr>
</tbody>
</table>
Figure 2. A genetic linkage map constructed for the 147 BC$_2$F$_4$ population that originated from the mating of $G$. max × $G$. soja. The map is comprised of 120 SSRs and one phenotypic marker. These markers coalesced into 20 linkage groups that were aligned, based on SSR marker homology, with the 20 known soybean linkage groups (Cregan et al., 1999). In each of the 20 linkage groups, marker names are given on the right, with the Haldane recombination distances (cM) separating those markers shown on the left.
CHAPTER FIVE
RESULTS
Polymorphism and Marker Segregation

The two parents were screened with 534 SSR markers. 284 (53%) polymorphic markers were identified. Of these, 130 markers at an average interval of 25 cM on the integrated soybean genetic map were used in the initial QTL discovery analysis. Of those 130 markers, 5 produced a multiple banding pattern that could not be reliably scored in the BIL screen, and 5 could not be scored due to weakly expressed banding patterns, or a lack of amplification. Ultimately, the 147 BILs were genotyped for 120 SSRs and one phenotypic marker.

The segregation ratio in the BC$_2$F$_4$ population for the two parental marker alleles at each marker locus was expected to be 85.94%M:3.13%H:10.93%S, or if heterozygotes are ignored, a ratio of 7.86M:1S was expected. The observed segregation data were evaluated by Chi-square test at $\alpha=0.05$ for each of the 121 markers and each of the 20 linkage groups (Table 2). For this population, the average genome proportion is 88.8% $G.\ max$ alleles and 10.6% $G.\ soja$ alleles. Significant skewing of genotypic classes was not observed for any markers used in this study ($\alpha=0.05$). None of the Chi-square values were significant ($\alpha=0.05$) for the 20 linkage groups.

Linkage Map

The constructed genetic map for this study (Figure 2) contained 121 markers, for a total length of 1506 cM (Haldane) or 50% of the soybean genome (estimated to be approximately 3000cM), with an average interval size of 12.5 cM. These 121 markers coalesced into 20 different linkage groups (LGs), leaving no markers in the unlinked class. The linkage groups were designated with names corresponding to the integrated public soybean genetic map (Cregan, et al., 1999).

Marker orders for the SSRs were identical to the published maps for nine of the LGs: B1, D1b, D2, G, H, I, M, N and O. The mapping order of the SSR markers for the other eleven LGs was in large part similar to the public soybean genetic map. Three kinds of differences, however, were observed: 1) the inversion of two markers closely linked to one another, 2) the inversion of two marker groups when a large gap exists between these two marker groups, 3) simultaneous occurrence of both inversions. Possible reasons for this are: first, the QTL mapping population size is not very large, second, an advanced backcross population is not the ideal population for making a genetic map; and third, the number of markers on each LG was too few to allow highly accurate mapping.
Seven gaps of 25cM or more existed between pairs of markers in this population (Figure 2). The maximum distance separating two markers was approximately 68 cM, which occurred in linkage group A2.

QTL mapping analysis was conducted using the orders and genetic distances from my linkage maps and also from Cregan’s linkage maps. Results were compared. The QTL identified did not change as a result of using different linkage maps. The QTL were associated with the same marker no matter which linkage map was used. The only difference was that the locations of the QTL changed together with the marker order. For example, in my linkage map, SSR marker Satt511 was mapped on top of linkage group A1 (0 cM). But by using Cregan’s linkage map order, Satt511 was located at the bottom of linkage group A1 (46.2 cM). A yield QTL was identified near Satt511 on linkage group A1 by using both linkage maps. The only difference was that the location of the QTL marked by Satt511 was changed simultaneously with the marker location of Satt511. Since QTL identification was marker order neutral, the SSR marker alignments in each of the 20 LGs equivalent to the soybean map of linkage groups published by Cregan et al. (1999) were used. By adjusting the marker orders for linkage groups A1, A2, B2, C1, C2, D1a, E, F, J, K, and L, the Map Manager software can be forced to calculate a new genetic distance when gene orders were changed.

**Field Data Analysis**

Performance of both the QTL mapping (A) and QTL verification (B) populations was evaluated in 2003 and 2004 at Lexington and Princeton, KY. In this study, these two years and two locations were treated as four environments in the analysis of variance procedure. In 2003 at Lexington, the plots of the QTL mapping (A) population were infected with tobacco ringspot virus producing bud blight infection. Extensive yield loss was observed on many plots, and a total yield loss was observed on 20 of the 147 plots. Other agronomic traits were seriously affected by the disease. A correlation analysis was conducted between environment 03Lex and the other three environments: 04 Lex, 03 Prin and 04Prin. Pearson correlation coefficients for yield between environment pairs (based on 147 BIL yield means of 2 reps for each environment) are shown in Table 3. Correlations between 03Lex and 03Prin, 04Lex and 04Prin were low, r=0.15, 0.17, and 0.05, respectively. Among them, only the correlation between 03Lex and 04Lex (r=0.17) was significant at P<0.05 level, but the correlation between these two environments is weak. The other three environment pairs showed strong positive correlations at
P<0.01 level: $r=0.51$ between 03Prin and 04Lex, $r=0.32$ between 03Prin and 04Prin, and $r=0.53$ between 04Prin and 04Lex (Table 3). I concluded the yield data from 03Lex were influenced by the severity of the soybean bud blight infection. Consequently, yield data from 03Lex were not used in the QTL discovery analysis.

Descriptive statistics for the 7499 and 147 BC$_2$F$_4$ lines were collected and data are shown in Table 4. The mean seed yield of the _G. max_ checks substantially exceeded that of the mean seed yield of the BC$_2$F$_4$ population. Mean yield of _G. max_ checks was 4311 kg/ha. Mean yield of lines was 3524 kg/ha ranging from 2219-4708 kg/ha. The range in values for the BILs bracketed the _G. max_ checks for all variables except 100 seed weight for which all BILs had smaller seeds (Table 4).

Analysis of variance was performed on each trait in the QTL analysis population at each of the three individual environments, and for the combined data (Table 5). There was significant (P<0.01) genetic variance among lines for yield, 100 seed weight, plant height and plant lodging in all single environments. Maturity set effects were not significant for any one of these four traits in 03Prin and 04Prin. Maturity set effects were significant for plant height in 04Lex, but not significant for seed yield, 100 seed weight and lodging. For the combined data, maturity set effects were not significant for seed yield, 100 seed weight, plant height and plant lodging. Environments had significant effects on 100 seed weight, plant height and lodging, but not on seed yield. Environment × line interactions were not significant for seed yield, but were significant for 100 seed weight, plant height and plant lodging.

Seed filling period and maturity were evaluated only at Lexington in 2003 and 2004; data were analyzed based on these two environments. Analysis of variance was performed on both single environment and combined data. To improve the effectiveness of the data, plots with severe bud blight disease were removed. Healthy plants were chosen from the middle of the row and seed filling period and maturity notes were taken only on these plants. A correlation analysis was conducted on the line means between environment 03Lex and 04Lex for seed filling period and maturity. Correlations between 03Lex and 04Lex was $r=0.34$ (P<0.0001) for seed filling period and $r=0.66$ (P<0.0001) for maturity. For both of the traits, 03Lex and 04Lex have strong positive correlations. I concluded that the seed filling period and maturity data from 03Lex could be used in the QTL discovery analysis.
The variance among lines for maturity was not significant in 03Lex, but highly significant in 04Lex. The variance among lines for seed filling period was highly significant for both 03Lex and 04Lex. For the combined data, variance among lines was significant for both maturity and seed filling period. The environment effects were not significant for maturity and seed filling period for the combined data. Maturity set effects were significant for maturity at $P<0.01$ level and significant for seed filling period at $P<0.05$ level for the combined data. Environment × line interactions were highly significant for both seed filling period and maturity.

**Trait Correlations**

Correlation values between seed yield and other traits based on means measured in the BC$_2$F$_4$ soybean population grouped within maturity sets 1, 2 and 3 were evaluated for statistical significance ($\alpha=0.05$) (Table 6). Agronomic traits maturity, seed filling period, and height were not significantly correlated with yield (Table 6). Only 100 seed weight had a significant positive correlation with yield within all three maturity sets. Lodging has a significant negative correlation with yield in two of the three maturity sets, and negative but non-significant in set 1.

Correlation values among traits (based on 147 BIL means, disregarding the division into maturity set, $\alpha=0.05$) are summarized in Table 7. Seed yield has a strong positive ($r=0.47$) correlation with 100 seed weight. Plant lodging was negatively ($r=-0.37$) correlated with yield. Weight per 100 seeds had a strong negative ($r=-0.56$) correlation with lodging. Seed filling period and 100 seed weight were positively correlated ($r=0.38$). Seed filling period and maturity had a positive correlation ($r=0.39$). Height and lodging were positively correlated ($r=0.55$). These results suggest that the higher 100 seed weight was one of the factors that resulted in higher yield of lines in this study. A longer seed filling period produced higher 100 seed weight. Tall plants exhibited a greater tendency to lodge. The longer seed filling period was due in part to longer days to maturity.

**QTL Discovery**

QTL were analyzed by simple linear regression (marker regression), simple interval mapping, and composite interval mapping after selecting background markers for genetic variance control. QTL analysis data for the 147 BC$_2$F$_4$ lines over combined environments (Tables 8, 9, 10) and for each single environment (Tables 11 and 12) are shown. QTL that were significantly associated ($P<0.0001$, LOD>3.0) with yield and yield component traits for both single and combined environment analysis by MR method are shown. QTL with the LOD>2.0
but less than 3.0 (Table 8) were used as background markers in the CIM analysis. For SIM and CIM QTL analysis with combined data, the 1000 permutation tests were used to define an empirical detection threshold at an experiment-wise P-value<0.01 (corresponding to a SIM mapping LOD>3.6 or a CIM LOD>3.8) and P<0.05(corresponding to a SIM and CIM LOD>3.0). For single environment SIM analysis, P<0.05 was used.

**Seed Yield**

As shown in Table 8, two QTL for seed yield were identified for the combined data by MR analysis. The *G. soja* allele at Satt511 on linkage group A1 was associated with increased seed yield (LOD=4.3)(Figure 3A), while the *G. soja* allele at the other QTL Satt285 on linkage group J decreased seed yield (LOD=2.2). The additive yield effect was 191 kg ha\(^{-1}\) upon the substitution of a *G. soja* allele for *G. max* allele for the QTL at Satt511 on LGA1. The additive effect was -168 kg ha\(^{-1}\) for the QTL at Satt285 on LGJ. The phenotypic variance accounted for by the QTL at Satt511 was 12%.

For the single environment QTL MR analysis (Table 11), seed yield QTL at Satt511 increased seed yield in all three environments, with LOD>2.0 in two environments and LOD=1.5 in the third environment. The additive yield effects of the QTL at Satt511 were 318, 129, and 185 kg ha\(^{-1}\), for the three environments. The phenotypic variances explained by the QTL at Satt511 were 15, 4, and 7%, for the three environments.

In the single environment QTL analysis, another seed yield QTL at Satt529 on LG-J was detected in one of the three environments 04Prin (LOD=3.4). The additive seed yield effect was 374 kg ha\(^{-1}\) with a \(R^2\) value of 10%. The yield QTL Satt529 (36.6cM) and Satt285 (0cM) on linkage group J have different locations on the chromosome, and different effects thus the expectation is that the QTL are different.

The results from the SIM analysis (Table 9) for the combined data indicated that a yield QTL mapping to LG-A1, between markers Satt050 and Satt511, with a mean position of 42 cM, and an additive yield effect of 223 kg ha\(^{-1}\) (LOD=4.4) (Figure 3A) upon the substitution of a *G.soja* allele for a *G.max* allele. The phenotypic variance accounted for by this QTL was 13%.

For the single environment SIM analysis (Table 12), positive additive yield effects were observed in all three environments for the substitution of the *G. soja* allele for the *G. max* allele at the location between Satt050 and Satt511. The average \(R^2\) of this QTL was 9.7% for the three environments.
The combined environment CIM results are presented in Table 10. The background markers used for yield were Satt285 and Satt050. The CIM-determined map position of seed yield QTL was very similar to the SIM determined map position on LG-A1. The LOD scores are very similar and both significant. The additive yield effect is 236 kg/ha for the CIM analysis compared to 223 kg/ha for the SIM analysis. The $R^2$ (13%) remained unchanged for seed yield.

As shown in Table 13, the yield increase with lines homozygous for the PI 245331 allele at Satt511 was significant in all three environments and for the combined data. For 04Lex, 03Prin and 04Prin, the yield increase was 7.5, 10.5 and 18.4%, with $P=0.0008$, 0.017 and <0.0001, respectively. The combined data showed a significant yield increase of 12% ($p<0.0001$) for lines homozygous for the PI 245331 allele over lines homozygous for the *Glycine max* ‘7499’ allele.

Allelic means and class differences for combined allelic types at markers Satt511 (significant in combined data) and Satt529 (significant in 04Prin) were calculated across the testing environments (Table 14). Four combined allelic types were identified: mm, ms, sm and ss. Lines with ‘mm’ genotypes are homozygous for ‘7499’ at both alleles. Lines with ‘ss’ genotypes are homozygous for PI 245331 at both alleles. Allelic means were 3469, 3506, 3765 and 4682 kg/ha, for classes mm, ms, sm and ss, respectively. The allelic class ‘ss’ had the maximum yield compared with the other three allelic types. Paired class differences were identified (Table 14). The class difference of mm-sm, mm-ss, ms-ss, and sm-ss were significantly different ($P<0.01$). Even though the QTL at Satt529 was identified only in 04Prin, the interaction of the two *G. soja* alleles was positive over the combined environments.

### 100-Seed Weight

Seven QTL for seed size were identified (LOD>2.0) from the combined data by MR (Table 8): the QTL at Satt304 on LG-B2 (LOD=3.3), QTL at Satt386 on LG-D2 (LOD=2.1), QTL at Satt324 and Satt594 on LG-G (LOD=2.2 and 2.8), QTL at Satt541 on LG-H (LOD=4.6), QTL at Satt549 and Satt339 on LG-N (LOD=2.8 and 2.6). The *G. soja* allele decreased seed weight for all QTL.

For the single environment MR analysis (Table 11): four QTL were detected in 04Prin, five QTL were detected in 03Prin, and 11 QTL were identified in 04Lex. QTL at Satt304 on LG-B2 and QTL at Satt541 on LG-H were repeatably identified in all three environments. The most significant QTL was Satt541 detected in 04Prin (LOD=4.5), 03Prin (LOD=4.4), and 04Lex.
(LOD=4.5), with a $R^2$ of 13% in all three environments. The additive effects for the 13 distinct QTL were negative as G. soja alleles at the 13 QTL decreased 100-seed weight.

In this study, QTL with LOD>3.0 and P<0.0001 were counted as significant in the MR analysis; QTL with LOD>3.6 and P<0.01 were counted as significant in the SIM analysis. The other QTL listed were used as background markers for the CIM analysis.

The SIM results of 100-seed weight for the combined data are presented in Table 9. Five QTL were detected (LOD>3.0) in LG-B2, D2, G, H, and N. As expected from the MR results, the most significant QTL was mapped to LG-H, between markers Satt192 and Satt541, with a mean position of 34 cM, and an additive effect -1.21 g (LOD=4.2) upon the substitution of a G. soja allele for G. max allele. The phenotypic variance accounted for by this QTL was 12%.

The SIM results of the single environment analysis are presented in Table 12. QTL on LG-B2, G and H were repeatable in all three environments.

The CIM-determined map position of 100-seed weight QTL was very similar to the SIM determined map position on LG B2, G, and N. But the significant QTL on the three linkage groups B2, G and N had CIM-generated LOD scores that were no longer significant. As expected from the SIM analysis, the most significant QTL was mapped to LG-H with LOD=5.0. The background markers used for the CIM analysis were Satt304, Satt324, Satt594, Satt192 and Satt549 (Table 10). The CIM percentage of trait variation ($R^2$) for 100-seed weight was less than that for SIM. CIM is good for detecting weaker QTL, but it can lessen the significance of the major QTL, because degrees of freedom are sacrificed to find lesser QTL.

The significant 100-seed weight QTL on LG-B2, G and N had CIM-generated LOD scores that were no longer significant. In contrast, the 100-seed weight QTL on LG-D2 and H remained significant (LOD=3.2 and 5.0) in the CIM analysis. But the QTL mean positions for these two QTL shifted rightward compared to the positions detected by SIM.

**Seed Filling Period**

Seed filling period (SFP) and maturity notes were taken only in Lexington, and SFP and MAT data were analyzed only for two environments (03Lex and 04Lex). Three SFP QTL were identified (LOD>2.0) for the combined data by MR analysis (Table 8) with only the QTL at Satt335 (Figure 3B) being significant (LOD=4.0).
The *G. soja* allele at Satt335 on LG-F increased SFP, while the *G. soja* allele at the other two QTL decreased SFP. The additive effect of the QTL at Satt335 was 1.13 days with $R^2$ of 12% over combined environments.

The MR results for the single environment QTL analysis identified three QTL in 03Lex but no QTL was found in 04Lex with LOD>3.0. The QTL at Satt335 on LG-F was found to be significant in one of the two environments, but positive additive effects were observed in both environments. The phenotypic variance accounted for by the QTL at Satt335 was 11% in 03Lex and 4% in 04Lex, respectively.

As expected from the MR results, a SFP QTL mapping to LG-F was detected by SIM analysis (Table 9 and Figure 3B), between markers Satt335 and Satt334, with a mean position of 73 cM, and an additive SFP effect of 1.28 days upon the substitution of a *G. soja* allele for *G. max* allele. The phenotypic variance accounted for by this QTL was 11%.

Results similar to the MR analysis were observed with the SIM analysis from single environments (Table 12). QTL between Satt335 and Satt334 on LG-F was found to be significant in 03Lex (LOD=4.5) and positive additive effects of this QTL were also observed in 04Lex. The phenotypic variance accounted for by the QTL was 13% for 03Lex and 2% for 04Lex, respectively. The significant seed filling period QTL on LG-F remained significant in the CIM analysis, with a similar map position and increased additive effects (Table 10). Three background markers were used for this CIM analysis. They were Satt334, Satt284 and Satt584 (Table 10).

As shown in Table 13, one (03Lex) out of two environments showed significant SFP increases in BC$_2$F$_4$ lines carrying the PI 245331 alleles at the SSR marker Satt335 locus as compared to BC$_2$F$_4$ lines homozygous for the *Glycine max* ‘7499’ alleles at the same locus. For the combined data, the SFP increases was 6.7% ($P=0.019$).

**Maturity**

Six maturity QTL were observed for the combined data by MR analysis with LOD>2.0 (Table 8). Two were significant (LOD>3.0), while the other four were used as background markers for CIM analysis. For the two significant QTL, QTL at Satt335 on LG-F has a positive additive effect of 1.4 days (LOD=3.3) with $R^2=10\%$. The other QTL at Satt584 on LG-N has a negative additive effect of -1.4 days (LOD=3.2) with $R^2=9\%$. Single environment MR analysis of maturity data identified the QTL at Satt335 on LG-F with positive additive effects in both
environments, but it was significant only in 04Lex (LOD=4.0). The phenotypic variance accounted for by this QTL was 12% in 04Lex (Table 11). Another QTL was identified in 04Lex on LG-N marked by Satt584 (LOD=3.1, P=0.0002) with a negative additive effect -1.7 and accounting for 9% of the phenotypic variance. The additive effects for QTL Satt584 on linkage group N were negative in both environments.

No significant QTL were identified by SIM analysis for the combined environments (Table 9). But as expected from the MR results, QTL at Satt359 on LG-B1, QTL at Satt335 on LG-F, and QTL at Satt331 on LG-O have positive additive effects of 1.3, 1.6 and 1.5 days, respectively. QTL at Satt584 on linkage group N has negative additive effects of -1.5 days (LOD=2.9) with $R^2$=9%.

Results similar to those from the MR analysis were identified by SIM for the single environment data. QTL at Satt335 on LG-F had positive additive effects in both environments, but it was significant only in 04Lex (LOD=3.0).

As shown in Table 13, in both environments, lines that were homozygous for the PI 245331 allele at the QTL Satt335 locus demonstrated significant increases (P<0.0001) in days to maturity over lines that were homozygous for the *Glycine max* '7499' allele. For the combined data, the increase in days to maturity was 10.4% (P<0.0001).

### Lodging

Six lodging QTL were identified for the combined data by MR analysis (Table 8). Three of the six QTL have LOD>3.0. For all of these QTL the *G. soja* allele increased lodging. The QTL with the largest effect was Satt284 on LG-L (LOD=5.2), with a maximum $R^2$ of 15%.

QTL at Satt284 on LG-L was repeatable in all three environments (Table 11). The phenotypic variance accounted for by this QTL was 19, 10 and 10%, respectively for the three environments.

Similar results were observed for both combined and single environment data by SIM analysis. QTL at Satt284 was identified as having the largest effect (LOD=5.8) (Table 9), and repeatable measurements were observed for all three environments (Table 12).

The significant lodging QTL on LG-L remained significant in the CIM analysis. Similar map position and additive effects were also observed. The $R^2$ is 15% for the CIM analysis compared to 16% with SIM (Table 10).
**Height**

Three height QTL were identified by MR for the combined data (Table 8). Two were on LG-L, the other one was on LG-F. None of them has LOD>3.0. They were used as background markers for CIM analysis. For all three QTL the *G. soja* allele increased height. Height QTL were detectable only in two of the three environments by MR analysis. The MR analysis for the single environment data was repeatable for LG-F at Satt144 and at Satt284 on LG-L. The QTL at Satt144 on LG-F was significant in both environments with LOD=3.2 for 04Prin and LOD=3.1 for 04Lex.

For the combined data a height QTL mapping to LG-L was detected by SIM analysis, between markers Satt652 and Satt284, with a mean position of 5 cM, and an additive height effect of 5.2 cm (LOD=3.6) upon the substitution of a *G. soja* allele for *G. max* allele. The phenotypic variance accounted for by this QTL was 11%. Similar results were observed for the single environment SIM analysis (Table 12). QTL between markers Satt652 and Satt284 was identified to be significant in both environments (LOD=3.0 and 3.3).

The significant height QTL (LOD=3.6) on LG-L had CIM-generated LOD scores that were no longer significant (LOD=2.6). The mean map position remained unchanged and additive effects were similar. The $R^2$ is 8% for CIM analysis compared to 11% for SIM (Table 10).

**QTL Confirmation**

To validate the QTL effects found in PI 245331, another 148 BC$_2$F$_4$ lines (population B) were used as a confirmation population. Population B was generated from the mating between recurrent parent *Glycine max* ‘7499’ and donor parent *Glycine soja* PI 245331. Population B and population A are related with one line in each population tracing back to the same BC$_2$ plant.

Three of four locations showed significant (P<0.25) yield increase with homozygosity for the PI 245331 allele at Satt511 (QTL1, Table 13, 15). The yield increase for 03Lex, 04Lex and 04Prin was 6.9, 7.6 and 12.8%, respectively (Table 13). Environment 03Prin had a yield decrease of -1.2% (P=0.70), but it was not significant. For the combined environments, lines that were homozygous for the PI 245331 allele at the QTL locus Satt511 on LG-A1 had a significant effect (P=0.08) (QTL1, Table 16). It demonstrated a 6.3% yield increase (P=0.037) over lines that were homozygous for the *Glycine max* ‘7499’ allele (Table 13).

Two out of four environments showed significant (P<0.25) yield increase with homozygosity for the PI 245331 allele with QTL marked by Satt529 on linkage group J (QTL2,
Table 15). For the combined environment, the QTL effect was significant (P<0.25) (QTL2, Table 16).

For seed yield in single environments (Table 15), only one of the four environments showed significant (03Lex, P=0.006) effects of QTL1 × QTL2. Set × QTL1 effect was significant in 03Prin (P=0.04) and was non-significant in 03Lex, 04Lex and 04Prin. Set × QTL2 effect was not significant in all four environments. None of the maturity set effects were significant in the four environments. For seed filling period in single environments, maturity set effect and set × QTL effect were not significant for both 03Lex and 04Lex. But for maturity effect, maturity set effects were significant (P<0.0001) in both 03Lex and 04Lex. Set × QTL effects were significant (P=0.01) only in 04Lex. For seed yield in combined environments (Table 16), QTL1 × QTL2 effect was significant (P=0.18). Non-significant effects were observed for maturity set, set × QTL1, set × QTL2. For seed filling period in combined environments (Table 16), maturity set and set × QTL effects were not significant. But for maturity effects in combined environments (Table 16), the maturity set effect was significant (P<0.0001).

Besides increased yield by QTL Satt511 on LG-A1, other phenotypic differences attributable to the presence of the PI 245331 allele were observed in the validation population. I identified one seed filling period QTL at Satt335 (LOD=4.0) on LG-F with an additive effect of +1 day. This QTL also provided a +1 day additive effect (LOD=3.3) on maturity. In the QTL validation population, this allele increased the seed filling period 1.5 days or 3.8% (P=0.015) for the combined environments (Table 13). Lines that were homozygous for the PI 245331 allele at Satt335 also demonstrated a 4.6% (P<0.0001) increase in days to maturity compared to lines that were homozygous for the Glycine max ‘7499’ allele (Table 13). For SFP, the allelic mean difference was significant in one of the two environments (Table 13). The QTL at Satt335 on LG-F increased the days to maturity in both environments 03Lex and 04Lex (P<0.0001) (Table 15). For the combined environments, the QTL marked by Satt335 on linkage group F significantly lengthened both SFP and days to maturity (P<0.0001) (Table 16).

Allelic means and class differences for combined allelic types at Satt511 and Satt529 were identified across the testing environments in population B (Table 14). Their allelic means of the four combined allelic types mm, ms, sm and ss, were 3055, 2734, 3209 and 3402 kg/ha, respectively. The allelic type ‘ss’ had the maximum yield compared with the other three allelic
types. The class difference of mm-sm, mm-ss, ms-ss, sm-ss were all significantly different (P<0.25).
Table 2. $X^2$ values for the fit to a 7.86:1 segregation ratio of *G. max* and *G. soja* alleles calculated for each marker, individual linkage groups, and all 121 markers in 147 BC$_2$F$_4$ BILs.

<table>
<thead>
<tr>
<th>LG† Marker</th>
<th>X$^2$‡</th>
<th>LG Marker</th>
<th>X$^2$</th>
<th>LG Marker</th>
<th>X$^2$</th>
<th>LG Marker</th>
<th>X$^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>A1</td>
<td>0.60</td>
<td>C2</td>
<td>0.44</td>
<td>F</td>
<td>0.23</td>
<td>K</td>
<td>0.01</td>
</tr>
<tr>
<td>Satt050</td>
<td>0.01</td>
<td>Satt281</td>
<td>0.56</td>
<td>Satt146</td>
<td>0.33</td>
<td>Satt137</td>
<td>0.30</td>
</tr>
<tr>
<td>Satt165</td>
<td>0.44</td>
<td>Satt363</td>
<td>1.22</td>
<td>Satt425</td>
<td>0.04</td>
<td>Satt337</td>
<td>0.09</td>
</tr>
<tr>
<td>Satt364</td>
<td>0.57</td>
<td>Satt371</td>
<td>0.22</td>
<td>Satt595</td>
<td>0.11</td>
<td>Satt001</td>
<td>0.04</td>
</tr>
<tr>
<td>Satt300</td>
<td>0.15</td>
<td>Satt227</td>
<td>0.03</td>
<td>Satt334</td>
<td>0.16</td>
<td>Satt273</td>
<td>0.02</td>
</tr>
<tr>
<td>Satt591</td>
<td>0.33</td>
<td>Satt681</td>
<td>0.16</td>
<td>Satt335</td>
<td>0.00</td>
<td>Satt260</td>
<td>0.18</td>
</tr>
<tr>
<td>Satt511</td>
<td>0.15</td>
<td>Satt100</td>
<td>1.85</td>
<td>Satt144</td>
<td>0.58</td>
<td>Satt196</td>
<td>0.01</td>
</tr>
<tr>
<td>A2</td>
<td>1.15</td>
<td>D1a</td>
<td>0.03</td>
<td>G</td>
<td>0.04</td>
<td>Satt588</td>
<td>0.96</td>
</tr>
<tr>
<td>Satt390</td>
<td>0.44</td>
<td>Satt184</td>
<td>0.16</td>
<td>Satt038</td>
<td>0.05</td>
<td>L</td>
<td>0.50</td>
</tr>
<tr>
<td>Satt315</td>
<td>0.84</td>
<td>Satt482</td>
<td>0.06</td>
<td>Satt324</td>
<td>0.00</td>
<td>Satt284</td>
<td>0.02</td>
</tr>
<tr>
<td>Satt508</td>
<td>2.70</td>
<td>Satt402</td>
<td>0.03</td>
<td>Satt594</td>
<td>0.14</td>
<td>Satt481</td>
<td>0.15</td>
</tr>
<tr>
<td>Satt429</td>
<td>0.29</td>
<td>Satt507</td>
<td>0.01</td>
<td>Satt199</td>
<td>0.02</td>
<td>Satt373</td>
<td>0.22</td>
</tr>
<tr>
<td>Satt133</td>
<td>2.32</td>
<td>D1b</td>
<td>1.28</td>
<td>Satt288</td>
<td>0.08</td>
<td>Satt652</td>
<td>0.10</td>
</tr>
<tr>
<td>B1</td>
<td>0.28</td>
<td>Satt157</td>
<td>0.38</td>
<td>Satt472</td>
<td>0.02</td>
<td>Satt664</td>
<td>0.24</td>
</tr>
<tr>
<td>Satt426</td>
<td>0.04</td>
<td>Satt428</td>
<td>0.47</td>
<td>H</td>
<td>0.50</td>
<td>Satt678</td>
<td>0.98</td>
</tr>
<tr>
<td>Satt251</td>
<td>0.00</td>
<td>Satt546</td>
<td>0.14</td>
<td>Satt192</td>
<td>0.15</td>
<td>Gene L1</td>
<td>0.42</td>
</tr>
<tr>
<td>Satt453</td>
<td>1.00</td>
<td>Satt271</td>
<td>0.04</td>
<td>Satt541</td>
<td>0.03</td>
<td>M</td>
<td>0.14</td>
</tr>
<tr>
<td>Satt597</td>
<td>0.10</td>
<td>Satt274</td>
<td>0.32</td>
<td>Satt568</td>
<td>0.22</td>
<td>Satt150</td>
<td>0.05</td>
</tr>
<tr>
<td>Satt359</td>
<td>0.10</td>
<td>Satt189</td>
<td>0.57</td>
<td>Satt666</td>
<td>0.05</td>
<td>Satt540</td>
<td>0.04</td>
</tr>
<tr>
<td>B2</td>
<td>0.12</td>
<td>D2</td>
<td>0.18</td>
<td>Satt142</td>
<td>0.50</td>
<td>Satt463</td>
<td>0.32</td>
</tr>
<tr>
<td>Satt577</td>
<td>0.33</td>
<td>Satt498</td>
<td>0.07</td>
<td>Satt181</td>
<td>0.30</td>
<td>Satt306</td>
<td>0.01</td>
</tr>
<tr>
<td>Satt063</td>
<td>0.03</td>
<td>Satt002</td>
<td>0.06</td>
<td>Satt434</td>
<td>0.09</td>
<td>Satt346</td>
<td>1.45</td>
</tr>
<tr>
<td>Satt168</td>
<td>0.02</td>
<td>Satt582</td>
<td>0.01</td>
<td>I</td>
<td>0.29</td>
<td>Satt551</td>
<td>0.02</td>
</tr>
<tr>
<td>Satt304</td>
<td>0.00</td>
<td>Satt311</td>
<td>0.33</td>
<td>Satt451</td>
<td>0.62</td>
<td>Satt336</td>
<td>0.15</td>
</tr>
<tr>
<td>Satt318</td>
<td>0.10</td>
<td>Satt301</td>
<td>0.02</td>
<td>Satt239</td>
<td>0.07</td>
<td>N</td>
<td>0.03</td>
</tr>
<tr>
<td>Satt726</td>
<td>0.20</td>
<td>Satt310</td>
<td>0.25</td>
<td>satt270</td>
<td>0.13</td>
<td>Satt152</td>
<td>0.00</td>
</tr>
<tr>
<td>Satt687</td>
<td>0.02</td>
<td>Satt386</td>
<td>0.39</td>
<td>Satt330</td>
<td>0.00</td>
<td>Satt584</td>
<td>0.05</td>
</tr>
<tr>
<td>C1</td>
<td>0.30</td>
<td>Satt461</td>
<td>0.00</td>
<td>Satt162</td>
<td>0.09</td>
<td>Satt387</td>
<td>0.00</td>
</tr>
<tr>
<td>Satt194</td>
<td>0.14</td>
<td>E</td>
<td>0.16</td>
<td>Satt440</td>
<td>0.05</td>
<td>Satt549</td>
<td>0.05</td>
</tr>
<tr>
<td>Satt399</td>
<td>0.02</td>
<td>Satt411</td>
<td>0.13</td>
<td>J</td>
<td>0.51</td>
<td>Satt339</td>
<td>0.00</td>
</tr>
<tr>
<td>Satt195</td>
<td>0.14</td>
<td>Satt185</td>
<td>0.10</td>
<td>Satt285</td>
<td>0.04</td>
<td>O</td>
<td>0.01</td>
</tr>
<tr>
<td>Satt578</td>
<td>0.02</td>
<td>Satt212</td>
<td>0.16</td>
<td>Satt596</td>
<td>0.03</td>
<td>Satt259</td>
<td>0.58</td>
</tr>
<tr>
<td>Satt718</td>
<td>0.08</td>
<td>Satt369</td>
<td>0.27</td>
<td>Satt414</td>
<td>0.00</td>
<td>Satt420</td>
<td>0.10</td>
</tr>
<tr>
<td>Satt164</td>
<td>0.10</td>
<td>Satt720</td>
<td>0.31</td>
<td>Satt529</td>
<td>0.68</td>
<td>Satt478</td>
<td>0.00</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Satt230</td>
<td>0.02</td>
<td>Satt431</td>
<td>0.44</td>
<td>Satt477</td>
<td>0.00</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Satt331</td>
<td>0.22</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Satt153</td>
<td>0.84</td>
</tr>
</tbody>
</table>

All markers $0.62$

$X^2_{(1, .05)} = 3.84$, $X^2_{(1, .01)} = 6.63$

LG†: linkage group indicated as a single letter or letter and numeral
Table 3. Pearson correlation coefficients for yield between environment pairs (based on 147 backcross inbred line yield means of 2 reps for each environment) in the BC$_2$F$_4$ mapping population that originated from the mating of *G. max* x *G. soja*.

<table>
<thead>
<tr>
<th>Environments</th>
<th>03Lex</th>
<th>03Prin</th>
<th>04Lex</th>
</tr>
</thead>
<tbody>
<tr>
<td>03Prin</td>
<td>0.15</td>
<td></td>
<td></td>
</tr>
<tr>
<td>04Lex</td>
<td>0.17*</td>
<td>0.51**</td>
<td></td>
</tr>
<tr>
<td>04Prin</td>
<td>0.05</td>
<td>0.32**</td>
<td>0.53**</td>
</tr>
</tbody>
</table>

*, **, significant effect at $\alpha=0.05$, or $\alpha=0.01$, respectively.
Table 4. Descriptive statistics for the *G. max* checks and 147 BC²F₄ lines that originated from a *G. max × G. soja* mating for seed yield, 100-seed weight, plant maturity, height, lodging, and seed filling period. The backcross inbred lines (BILs) and *G. max* checks were evaluated in 2003 and 2004 field trials at two locations and with two replicates (8 observations per BIL). Seed yield, 100 seed weight, height and lodging were evaluated in 3 environments. Maturity and seed filling period were evaluated in 2 environments.

<table>
<thead>
<tr>
<th>Traits</th>
<th>SY§</th>
<th>100-SW£</th>
<th>MAT#</th>
<th>HT†</th>
<th>LDG‡</th>
<th>SFP§</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>units</strong></td>
<td>kg/ha</td>
<td>g</td>
<td>days</td>
<td>cm</td>
<td>score</td>
<td>days</td>
</tr>
<tr>
<td><strong>G. max checks</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean</td>
<td>4310.7</td>
<td>15.9</td>
<td>22.9</td>
<td>93.7</td>
<td>1.35</td>
<td>42.1</td>
</tr>
<tr>
<td>SD¥</td>
<td>679.2</td>
<td>1.4</td>
<td>2.3</td>
<td>7.6</td>
<td>0.5</td>
<td>3.6</td>
</tr>
<tr>
<td><strong>BILs</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean</td>
<td>3523.9</td>
<td>10.8</td>
<td>27.3</td>
<td>92.6</td>
<td>2.5</td>
<td>37.4</td>
</tr>
<tr>
<td>Minimum</td>
<td>2219.3</td>
<td>5.6</td>
<td>18.5</td>
<td>74.1</td>
<td>1</td>
<td>30</td>
</tr>
<tr>
<td>Maximum</td>
<td>4707.5</td>
<td>15.5</td>
<td>38</td>
<td>125.1</td>
<td>4.7</td>
<td>46</td>
</tr>
<tr>
<td>SD¥</td>
<td>376.6</td>
<td>1.8</td>
<td>3.2</td>
<td>8.2</td>
<td>0.8</td>
<td>2.4</td>
</tr>
</tbody>
</table>

§ Seed yield, kg/ha, was measured within each line
£ 100 seed weight, g adjusted to a 13% seed water content
# days from 30 August to the date when 95% of the pods had reached a final pod color
† height, distance from the soil to the tip of the main stem at maturity
‡ lodging, 1= all plants standing erect, 5= all plants prostrate
§ seed filling period, number of days from the date of R5 to the date of R7
¥ standard deviation
Table 5. Significance levels from the analysis of variance performed on each trait in population A at each environment, and for the combined data.

<table>
<thead>
<tr>
<th>Environment</th>
<th>Source of variation</th>
<th>Traits</th>
<th>Environment</th>
<th>Traits</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>SY§</td>
<td>100-SW₤</td>
<td></td>
</tr>
<tr>
<td>03Prin</td>
<td>Set ††</td>
<td>ns</td>
<td>ns</td>
<td>ns</td>
</tr>
<tr>
<td></td>
<td>Rep</td>
<td>ns</td>
<td>ns</td>
<td>ns</td>
</tr>
<tr>
<td></td>
<td>Rep*set</td>
<td>ns</td>
<td>*</td>
<td>ns</td>
</tr>
<tr>
<td></td>
<td>Line ‡‡ (set)</td>
<td>**</td>
<td>**</td>
<td>**</td>
</tr>
<tr>
<td>04Lex</td>
<td>Set</td>
<td>ns</td>
<td>ns</td>
<td>*</td>
</tr>
<tr>
<td></td>
<td>Rep</td>
<td>ns</td>
<td>*</td>
<td>ns</td>
</tr>
<tr>
<td></td>
<td>Rep*set</td>
<td>**</td>
<td>*</td>
<td>ns</td>
</tr>
<tr>
<td></td>
<td>Line(set)</td>
<td>**</td>
<td>**</td>
<td>**</td>
</tr>
<tr>
<td>04Prin</td>
<td>Set</td>
<td>ns</td>
<td>ns</td>
<td>ns</td>
</tr>
<tr>
<td></td>
<td>Rep</td>
<td>ns</td>
<td>ns</td>
<td>ns</td>
</tr>
<tr>
<td></td>
<td>Rep*set</td>
<td>**</td>
<td>**</td>
<td>**</td>
</tr>
<tr>
<td></td>
<td>Line(set)</td>
<td>**</td>
<td>**</td>
<td>**</td>
</tr>
<tr>
<td>Combined</td>
<td>Env</td>
<td>ns</td>
<td>*</td>
<td>**</td>
</tr>
<tr>
<td></td>
<td>Set</td>
<td>ns</td>
<td>ns</td>
<td>ns</td>
</tr>
<tr>
<td></td>
<td>Set*env</td>
<td>ns</td>
<td>ns</td>
<td>ns</td>
</tr>
<tr>
<td></td>
<td>Rep(env)</td>
<td>ns</td>
<td>*</td>
<td>ns</td>
</tr>
<tr>
<td></td>
<td>Set*rep(env)</td>
<td>**</td>
<td>**</td>
<td>**</td>
</tr>
<tr>
<td></td>
<td>Line(set)</td>
<td>**</td>
<td>**</td>
<td>**</td>
</tr>
<tr>
<td>Combined</td>
<td>Env*line(set)</td>
<td>ns</td>
<td>**</td>
<td>**</td>
</tr>
</tbody>
</table>

* , ** significant effect at α=0.05, or α=0.01, respectively
§ seed yield, kg/ha, was measured within each line
₤ 100 seed weight, g adjusted to a 13% seed water content
# days from 30 August to the date when 95% of the pods had reached a final pod color
† height, distance from the soil to the tip of the main stem at maturity
‡ lodging, 1= all plants standing erect, 5= all plants prostrate
§ seed filling period, number of days from the date of R5 to the date of R7
†† BC₂F₄ lines grouped in sets 1, 2, and 3 according to maturity
‡‡ BC₂F₄ derived lines
Table 6. Pearson correlation coefficients between seed yield and other traits based on means measured in the BC$_2$F$_4$ mapping population grouped in sets 1, 2, and 3 according to maturity.

<table>
<thead>
<tr>
<th>Trait</th>
<th>Sets§</th>
<th>1</th>
<th>2</th>
<th>3</th>
</tr>
</thead>
<tbody>
<tr>
<td>100-SW§</td>
<td>0.31*</td>
<td>0.59**</td>
<td>0.56**</td>
<td></td>
</tr>
<tr>
<td>MAT#</td>
<td>-0.07</td>
<td>-0.12</td>
<td>-0.17</td>
<td></td>
</tr>
<tr>
<td>SFP$</td>
<td>0.10</td>
<td>0.26</td>
<td>0.16</td>
<td></td>
</tr>
<tr>
<td>HT†</td>
<td>-0.16</td>
<td>0.12</td>
<td>-0.07</td>
<td></td>
</tr>
<tr>
<td>LDG‡</td>
<td>-0.20</td>
<td>-0.58**</td>
<td>-0.38*</td>
<td></td>
</tr>
</tbody>
</table>

* **, significant effect at \( \alpha = 0.05 \), or \( \alpha = 0.01 \), respectively
§ BC$_2$F$_4$ lines grouped in sets 1, 2, and 3 according to maturity
₤ 100 seed weight, g adjusted to a 13% seed water content
# days from 30 August to the date when 95% of the pods had reached a final pod color
† height, distance from the soil to the tip of the main stem at maturity
‡ lodging, 1= all plants standing erect, 5= all plants prostrate
$ seed filling period, number of days from the date of R5 to the date of R7
Table 7. Pearson correlation coefficients between trait pairs (based on 147 individual environment BIL means) in the BC$_2$F$_4$ mapping population that originated from the mating of *G. max* × *G. soja*.

<table>
<thead>
<tr>
<th>Trait</th>
<th>SY$§$</th>
<th>HT$†$</th>
<th>LDG$‡$</th>
<th>MAT$#$</th>
<th>100-SW$₤$</th>
</tr>
</thead>
<tbody>
<tr>
<td>HT$†$</td>
<td>-0.05</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>LDG$‡$</td>
<td>-0.37**</td>
<td>0.55**</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MAT$#$</td>
<td>-0.11</td>
<td>0.25**</td>
<td>0.23**</td>
<td></td>
<td></td>
</tr>
<tr>
<td>100-SW$₤$</td>
<td>0.47**</td>
<td>-0.25**</td>
<td>-0.56**</td>
<td>-0.14</td>
<td></td>
</tr>
<tr>
<td>SFP$§$</td>
<td>0.15</td>
<td>-0.12</td>
<td>-0.27**</td>
<td>0.39**</td>
<td>0.38**</td>
</tr>
</tbody>
</table>

$*$ ***, significant effect at $\alpha=0.05$, or $\alpha=0.01$, respectively

$§$ seed yield, kg/ha, was measured within each line

$₤$ 100 seed weight, g adjusted to a 13% seed water content

$#$ days from 30 August to the date when 95% of the pods had reached a final pod color

$†$ height, distance from the soil to the tip of the main stem at maturity

$‡$ lodging, 1= all plants standing erect, 5= all plants prostrate

$§$ seed filling period, number of days from the date of R5 to the date of R7
Table 8. Markers of the indicated soybean linkage group and map position that were putatively linked to QTL based on fitting the phenotypic data of each trait to a marker regression model for the combined environments. The phenotypic data were the means of 147 BC$_2$F$_4$ lines that segregated about 7.86MM: 1SS for each of 121 markers.

<table>
<thead>
<tr>
<th>Trait</th>
<th>LG††</th>
<th>Marker</th>
<th>Position(cM)</th>
<th>LOD$^s$</th>
<th>P value</th>
<th>A##</th>
<th>R$^2$(%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>SY§ (Kg/ha)</td>
<td>A1</td>
<td>Satt511</td>
<td>46.2</td>
<td>4.3*</td>
<td>0.00001</td>
<td>190.60</td>
<td>12</td>
</tr>
<tr>
<td>J</td>
<td>Satt285</td>
<td>0</td>
<td>2.2</td>
<td>0.001</td>
<td>-168.3</td>
<td>7</td>
<td></td>
</tr>
<tr>
<td>100-SW§ (g)</td>
<td>B2</td>
<td>Satt304</td>
<td>32.5</td>
<td>3.3*</td>
<td>0.00009</td>
<td>-0.87</td>
<td>7</td>
</tr>
<tr>
<td>D2</td>
<td>Satt386</td>
<td>86.9</td>
<td>2.1</td>
<td>0.002</td>
<td>-0.63</td>
<td>7</td>
<td></td>
</tr>
<tr>
<td>G</td>
<td>Satt324</td>
<td>26.1</td>
<td>2.2</td>
<td>0.002</td>
<td>-0.69</td>
<td>7</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Satt594</td>
<td>38.9</td>
<td>2.8</td>
<td>0.0003</td>
<td>-0.86</td>
<td>8</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Satt541</td>
<td>38.5</td>
<td>4.6*</td>
<td>0.00000</td>
<td>-0.93</td>
<td>14</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Satt549</td>
<td>37.7</td>
<td>2.8</td>
<td>0.0003</td>
<td>-0.76</td>
<td>8</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Satt339</td>
<td>45.6</td>
<td>2.6</td>
<td>0.0005</td>
<td>-0.85</td>
<td>8</td>
<td></td>
</tr>
<tr>
<td>LDG‡ (score)</td>
<td>G</td>
<td>Satt594</td>
<td>38.9</td>
<td>2.1</td>
<td>0.002</td>
<td>0.33</td>
<td>6</td>
</tr>
<tr>
<td>I</td>
<td>Satt330</td>
<td>61.3</td>
<td>3.3*</td>
<td>0.00009</td>
<td>0.38</td>
<td>10</td>
<td></td>
</tr>
<tr>
<td>L</td>
<td>Satt284</td>
<td>11.4</td>
<td>5.2*</td>
<td>0.00000</td>
<td>0.48</td>
<td>15</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Satt652</td>
<td>0.0</td>
<td>2.8</td>
<td>0.0003</td>
<td>0.41</td>
<td>8</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Satt678</td>
<td>38.6</td>
<td>2.1</td>
<td>0.002</td>
<td>0.42</td>
<td>6</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Satt549</td>
<td>37.7</td>
<td>3.5*</td>
<td>0.00006</td>
<td>0.37</td>
<td>10</td>
<td></td>
</tr>
<tr>
<td>SFP$^s$ (days)</td>
<td>F</td>
<td>Satt335</td>
<td>72.6</td>
<td>4.0*</td>
<td>0.00001</td>
<td>1.13</td>
<td>12</td>
</tr>
<tr>
<td>L</td>
<td>Satt284</td>
<td>11.4</td>
<td>2.4</td>
<td>0.0009</td>
<td>-0.99</td>
<td>7</td>
<td></td>
</tr>
<tr>
<td>N</td>
<td>Satt584</td>
<td>7.1</td>
<td>2.9</td>
<td>0.0003</td>
<td>-1.01</td>
<td>9</td>
<td></td>
</tr>
<tr>
<td>MAT## (days)</td>
<td>A1</td>
<td>Satt364</td>
<td>3.5</td>
<td>2.3</td>
<td>0.001</td>
<td>-1.87</td>
<td>7</td>
</tr>
<tr>
<td>B1</td>
<td>Satt359</td>
<td>36.5</td>
<td>2.2</td>
<td>0.002</td>
<td>1.15</td>
<td>7</td>
<td></td>
</tr>
<tr>
<td>F</td>
<td>Satt335</td>
<td>74.1</td>
<td>3.3*</td>
<td>0.0001</td>
<td>1.37</td>
<td>10</td>
<td></td>
</tr>
<tr>
<td>J</td>
<td>Satt529</td>
<td>36.6</td>
<td>2.1</td>
<td>0.002</td>
<td>-1.69</td>
<td>6</td>
<td></td>
</tr>
<tr>
<td>N</td>
<td>Satt584</td>
<td>7.1</td>
<td>3.2*</td>
<td>0.0001</td>
<td>-1.40</td>
<td>9</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Satt339</td>
<td>45.6</td>
<td>2.0</td>
<td>0.002</td>
<td>-1.31</td>
<td>6</td>
<td></td>
</tr>
<tr>
<td>O</td>
<td>Satt331</td>
<td>49.7</td>
<td>2.6</td>
<td>0.0006</td>
<td>1.57</td>
<td>8</td>
<td></td>
</tr>
<tr>
<td>HT† (cm)</td>
<td>F</td>
<td>Satt144</td>
<td>91.9</td>
<td>2.9</td>
<td>0.0003</td>
<td>5.78</td>
<td>9</td>
</tr>
<tr>
<td>L</td>
<td>Satt652</td>
<td>0.0</td>
<td>2.6</td>
<td>0.0006</td>
<td>4.02</td>
<td>8</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Satt284</td>
<td>11.4</td>
<td>2.7</td>
<td>0.0005</td>
<td>3.55</td>
<td>8</td>
<td></td>
</tr>
</tbody>
</table>

*significant effect at $\alpha=0.0001$

§seed yield, kg/ha, was measured within each line

£100 seed weight, g adjusted to a 13% seed water content

#days from 30 August to the date when 95% of the pods had reached a final pod color

†height, distance from the soil to the tip of the main stem at maturity

‡lodging, 1= all plants standing erect, 5= all plants prostrate

§seed filling period, number of days from the date of R5 to the date of R7

††Linkage group

$^s$Logarithm of odds score, the LOD score significance criterion was 3 for MR analysis

##MR estimated additive effect, the additive effect (sign and magnitude) is that calculated for the substitution of a single $G. soja$ (s) allele for a $G. max$ (m) allele at the indicated marker locus

---

63
Table 9. Putative QTL identified by the simple interval mapping (SIM) method for the combined environment. The phenotypic data were the means of the 147 BC$_2$F$_4$ that segregated about 7.86MM:1SS for each of 121 markers. The mean map position of the significant QTL was established by the map position of the peak LOD score exceeding 3 in the interval between the two indicated flanking markers. A one-LOD fall-off (from the QTL peak) method was used to estimate the left and right flanking map positions of a confidence interval surrounding the mean QTL map position.

<table>
<thead>
<tr>
<th>Trait</th>
<th>LG$^{††}$</th>
<th>Marker on Left</th>
<th>QTL left flank (cM)</th>
<th>QTL mean position (cM)</th>
<th>Marker on Right</th>
<th>LOD$^{**}$</th>
<th>A$^{##}$</th>
<th>R$^2$ (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>SY$^§$ (Kg/ha)</td>
<td>A1</td>
<td>Satt050</td>
<td>35</td>
<td>42</td>
<td>45</td>
<td>Satt511</td>
<td>4.4**</td>
<td>223.45</td>
</tr>
<tr>
<td>100-SW$^\xi$ (g)</td>
<td>B2</td>
<td>Satt577</td>
<td>6</td>
<td>12</td>
<td>22</td>
<td>Satt168</td>
<td>3.0*</td>
<td>-1.45</td>
</tr>
<tr>
<td></td>
<td>D2</td>
<td>Satt002</td>
<td>0</td>
<td>17</td>
<td>38</td>
<td>Satt582</td>
<td>3.0*</td>
<td>-0.93</td>
</tr>
<tr>
<td></td>
<td>G</td>
<td>Satt324</td>
<td>30</td>
<td>35</td>
<td>42</td>
<td>Satt594</td>
<td>3.5*</td>
<td>-1.19</td>
</tr>
<tr>
<td></td>
<td>H</td>
<td>Satt192</td>
<td>28</td>
<td>34</td>
<td>50</td>
<td>Satt541</td>
<td>4.2**</td>
<td>-1.21</td>
</tr>
<tr>
<td></td>
<td>I</td>
<td>Satt239</td>
<td>13</td>
<td>28</td>
<td>38</td>
<td>Satt270</td>
<td>2.9</td>
<td>-0.97</td>
</tr>
<tr>
<td></td>
<td>N</td>
<td>Satt549</td>
<td>36</td>
<td>41</td>
<td>45</td>
<td>Satt339</td>
<td>3.8**</td>
<td>-1.03</td>
</tr>
<tr>
<td>LDG$^\dagger$ (score)</td>
<td>I</td>
<td>Satt330</td>
<td>53</td>
<td>61</td>
<td>67</td>
<td>Satt162</td>
<td>2.6</td>
<td>0.36</td>
</tr>
<tr>
<td></td>
<td>L</td>
<td>Satt652</td>
<td>4</td>
<td>7</td>
<td>19</td>
<td>Satt284</td>
<td>5.8**</td>
<td>0.61</td>
</tr>
<tr>
<td></td>
<td>N</td>
<td>Satt387</td>
<td>29</td>
<td>36</td>
<td>43</td>
<td>Satt549</td>
<td>2.6</td>
<td>0.35</td>
</tr>
<tr>
<td>SFP$^$$ (days)</td>
<td>F</td>
<td>Satt335</td>
<td>70</td>
<td>73</td>
<td>76</td>
<td>Satt334</td>
<td>3.8**</td>
<td>1.28</td>
</tr>
<tr>
<td>MAT$^{#}$ (days)</td>
<td>B1</td>
<td>Satt359</td>
<td>31</td>
<td>37</td>
<td>47</td>
<td>Satt453</td>
<td>2.7</td>
<td>1.32</td>
</tr>
<tr>
<td></td>
<td>F</td>
<td>Satt335</td>
<td>68</td>
<td>75</td>
<td>80</td>
<td>Satt334</td>
<td>2.7</td>
<td>1.61</td>
</tr>
<tr>
<td></td>
<td>N</td>
<td>Satt584</td>
<td>0</td>
<td>8</td>
<td>16</td>
<td>Satt387</td>
<td>2.9</td>
<td>-1.46</td>
</tr>
<tr>
<td></td>
<td>O</td>
<td>Satt477</td>
<td>41</td>
<td>49</td>
<td>57</td>
<td>Satt331</td>
<td>2.0</td>
<td>1.53</td>
</tr>
<tr>
<td>HT$^#$ (cm)</td>
<td>L</td>
<td>Satt652</td>
<td>0</td>
<td>5</td>
<td>14</td>
<td>Satt284</td>
<td>3.6**</td>
<td>5.20</td>
</tr>
</tbody>
</table>

* *, ** significant effect at $\alpha=0.05$, or $\alpha=0.01$, respectively

$^§$ seed yield, kg/ha, was measured within each line

$^\xi$ 100 seed weight, g adjusted to a 13% seed water content

$^#$ days from 30 August to the date when 95% of the pods had reached a final pod color

$^\dagger$ height, distance from the soil to the tip of the main stem at maturity

$^\$ lodging, 1= all plants standing erect, 5= all plants prostrate

$^\$$ seed filling period, number of days from the date of R5 to the date of R7

$^{##}$ SIM estimated additive effect, the additive effect (sign and magnitude) was that calculated for the substitution of a single $G$. soja allele for a $G$. max allele at the indicated QTL peak position

$^{***}$Logarithm of odds score, the LOD score significance criterion was LOD>3.6 corresponding to P<0.01, and LOD>3.0 corresponding to P<0.05.

$^{††}$Linkage group
Table 10. Putative QTL identified by the composite interval mapping (CIM) method for the combined environment, using as background markers those listed in table 8. The phenotypic data were the means of the 147 BC$_2$F$_4$ lines for each trait. These BILs segregated about 7.86MM:1SS for each of 121 markers. The map position of a significant QTL was established by the map position of the peak LOD score in the interval between the two indicated flanking markers. A one-LOD fall-off (from the QTL peak) method was used to estimate the left and right flanking map positions of a confidence interval surrounding the mean QTL map position.

<table>
<thead>
<tr>
<th>Trait</th>
<th>LG$^††$</th>
<th>Marker on Left</th>
<th>QTL left flank (cM)</th>
<th>QTL mean position (cM)</th>
<th>Marker on Right</th>
<th>LOD$^*$</th>
<th>A$^{**}$</th>
<th>R$^2$ (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>SY$^§$(Kg/ha)</td>
<td>A1</td>
<td>Satt050</td>
<td>34</td>
<td>41</td>
<td>45</td>
<td>Satt511</td>
<td>4.7**</td>
<td>235.5</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Background markers: Satt285, Satt050</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>100-SW$^\£$(g)</td>
<td>D2</td>
<td>Satt582</td>
<td>24</td>
<td>31</td>
<td>38</td>
<td>Satt461</td>
<td>3.2*</td>
<td>-0.91</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Background markers: Satt304, Satt594, Satt541, Satt549, Satt002</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>H</td>
<td>Satt541</td>
<td>40</td>
<td>47</td>
<td>52</td>
<td>Satt142</td>
<td>5.0**</td>
<td>-1.53</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Background markers: Satt304, Satt324, Satt594, Satt192, Satt549</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>LDG$^‡$(score)</td>
<td>L</td>
<td>Satt652</td>
<td>5</td>
<td>8</td>
<td>22</td>
<td>Satt284</td>
<td>6.5**</td>
<td>0.58</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Background markers: Satt594, Satt330, Satt652, Satt549</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SFP$^$$ (days)</td>
<td>F</td>
<td>Satt595</td>
<td>59</td>
<td>71</td>
<td>76</td>
<td>Satt335</td>
<td>4.8**</td>
<td>1.42</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Background markers: Satt334, Satt284, Satt584</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*, ** significant effect at $\alpha=0.05$, or $\alpha=0.01$, respectively

$^§$seed yield, kg/ha, was measured within each line

$^\£$100 seed weight, g adjusted to a 13% seed water content

$^‡$lodging, 1= all plants standing erect, 5= all plants prostrate

$^\$$seed filling period, number of days from the date of R5 to the date of R7

$^{**}$CIM estimated additive effect, the additive effect (sign and magnitude) was that calculated for the substitution of a single $G. soja$ allele for a $G. max$ allele at the indicated QTL peak position

$^*$Logarithm of odds score, the LOD score significance criterion was LOD>3.8 corresponding to $P<0.01$, and LOD>3.0 corresponding to $P<0.05$.

$^††$Linkage group
Table 11. Markers of the indicated soybean linkage group and map position that were putatively linked to QTL based on fitting the phenotypic data of population A of each trait to a marker regression model for the single environments.

<table>
<thead>
<tr>
<th>Trait</th>
<th>Env</th>
<th>LG\textsuperscript{††}</th>
<th>Marker</th>
<th>Position</th>
<th>LOD\textsuperscript{‡}</th>
<th>P value</th>
<th>A\textsuperscript{***}</th>
<th>R\textsuperscript{‡} (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>SY\textsuperscript{§} (Kg/ha)</td>
<td>04 Lex</td>
<td>A1</td>
<td>Satt511</td>
<td>46.2</td>
<td>2.5</td>
<td>0.0007</td>
<td>184.65</td>
<td>7</td>
</tr>
<tr>
<td></td>
<td></td>
<td>J</td>
<td>Satt285</td>
<td>0</td>
<td>2.8</td>
<td>0.0003</td>
<td>-235.20</td>
<td>9</td>
</tr>
<tr>
<td>03 Prin</td>
<td></td>
<td>A1</td>
<td>Satt511</td>
<td>46.2</td>
<td>1.5</td>
<td>0.009</td>
<td>129.40</td>
<td>4</td>
</tr>
<tr>
<td>04 Prin</td>
<td></td>
<td>A1</td>
<td>Satt511</td>
<td>46.2</td>
<td>5.3*</td>
<td>0.00000</td>
<td>317.95</td>
<td>15</td>
</tr>
<tr>
<td></td>
<td></td>
<td>J</td>
<td>Satt529</td>
<td>36.6</td>
<td>3.4*</td>
<td>0.00008</td>
<td>374.40</td>
<td>10</td>
</tr>
<tr>
<td>100-SW\textsuperscript{£} (g)</td>
<td>04 Lex</td>
<td>B2</td>
<td>Satt304</td>
<td>32.5</td>
<td>2.2</td>
<td>0.001</td>
<td>-0.78</td>
<td>7</td>
</tr>
<tr>
<td></td>
<td></td>
<td>D2</td>
<td>Satt498</td>
<td>0</td>
<td>2.2</td>
<td>0.001</td>
<td>-0.78</td>
<td>7</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Satt002</td>
<td>12.2</td>
<td>2.5</td>
<td>0.0006</td>
<td>-0.75</td>
<td>8</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Satt582</td>
<td>24.2</td>
<td>2.4</td>
<td>0.0009</td>
<td>-0.77</td>
<td>7</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Satt461</td>
<td>39.8</td>
<td>2.3</td>
<td>0.001</td>
<td>-0.82</td>
<td>7</td>
</tr>
<tr>
<td></td>
<td></td>
<td>G</td>
<td>Satt594</td>
<td>38.9</td>
<td>2.7</td>
<td>0.0005</td>
<td>-0.91</td>
<td>8</td>
</tr>
<tr>
<td></td>
<td></td>
<td>H</td>
<td>Satt541</td>
<td>38.5</td>
<td>4.5*</td>
<td>0.0001</td>
<td>-1.00</td>
<td>13</td>
</tr>
<tr>
<td></td>
<td></td>
<td>L</td>
<td>Satt284</td>
<td>11.4</td>
<td>2.3</td>
<td>0.001</td>
<td>-0.81</td>
<td>7</td>
</tr>
<tr>
<td></td>
<td></td>
<td>M</td>
<td>Satt551</td>
<td>50.6</td>
<td>2.4</td>
<td>0.0008</td>
<td>-0.90</td>
<td>7</td>
</tr>
<tr>
<td></td>
<td></td>
<td>N</td>
<td>Satt549</td>
<td>37.7</td>
<td>2.9</td>
<td>0.0002</td>
<td>-0.85</td>
<td>9</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Satt339</td>
<td>45.6</td>
<td>2.1</td>
<td>0.002</td>
<td>-0.84</td>
<td>7</td>
</tr>
<tr>
<td>03 Prin</td>
<td></td>
<td>B1</td>
<td>Satt426</td>
<td>0</td>
<td>2.4</td>
<td>0.0008</td>
<td>-0.82</td>
<td>8</td>
</tr>
<tr>
<td></td>
<td></td>
<td>B2</td>
<td>Satt304</td>
<td>32.5</td>
<td>2.8</td>
<td>0.0003</td>
<td>-0.83</td>
<td>9</td>
</tr>
<tr>
<td></td>
<td></td>
<td>G</td>
<td>Satt324</td>
<td>26.1</td>
<td>2.3</td>
<td>0.001</td>
<td>-0.73</td>
<td>7</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Satt594</td>
<td>38.9</td>
<td>2.4</td>
<td>0.0008</td>
<td>-0.83</td>
<td>7</td>
</tr>
<tr>
<td></td>
<td></td>
<td>H</td>
<td>Satt541</td>
<td>38.5</td>
<td>4.4*</td>
<td>0.00001</td>
<td>-0.94</td>
<td>13</td>
</tr>
<tr>
<td>04 Prin</td>
<td></td>
<td>B2</td>
<td>Satt304</td>
<td>32.5</td>
<td>2.6</td>
<td>0.0005</td>
<td>-0.75</td>
<td>8</td>
</tr>
<tr>
<td></td>
<td></td>
<td>G</td>
<td>Satt594</td>
<td>38.9</td>
<td>3.3*</td>
<td>0.0001</td>
<td>-0.89</td>
<td>10</td>
</tr>
<tr>
<td></td>
<td></td>
<td>H</td>
<td>Satt541</td>
<td>38.5</td>
<td>4.5</td>
<td>0.00001</td>
<td>-0.88</td>
<td>13</td>
</tr>
<tr>
<td></td>
<td></td>
<td>N</td>
<td>Satt549</td>
<td>37.7</td>
<td>2.1</td>
<td>0.002</td>
<td>-0.64</td>
<td>6</td>
</tr>
<tr>
<td>LDG\textsuperscript{‡} (score)</td>
<td>04 Lex</td>
<td>A2</td>
<td>Satt315</td>
<td>21.3</td>
<td>2.2</td>
<td>0.001</td>
<td>0.52</td>
<td>7</td>
</tr>
<tr>
<td></td>
<td></td>
<td>I</td>
<td>Satt330</td>
<td>61.3</td>
<td>2.2</td>
<td>0.001</td>
<td>0.35</td>
<td>7</td>
</tr>
<tr>
<td></td>
<td></td>
<td>L</td>
<td>Satt284</td>
<td>11.4</td>
<td>3.4*</td>
<td>0.00007</td>
<td>0.44</td>
<td>10</td>
</tr>
<tr>
<td></td>
<td></td>
<td>N</td>
<td>Satt549</td>
<td>37.7</td>
<td>2.97</td>
<td>0.0002</td>
<td>0.39</td>
<td>9</td>
</tr>
<tr>
<td>03 Prin</td>
<td></td>
<td>G</td>
<td>Satt594</td>
<td>38.9</td>
<td>2.95</td>
<td>0.0002</td>
<td>0.38</td>
<td>9</td>
</tr>
<tr>
<td></td>
<td></td>
<td>I</td>
<td>Satt330</td>
<td>61.3</td>
<td>3.4*</td>
<td>0.00007</td>
<td>0.38</td>
<td>10</td>
</tr>
<tr>
<td></td>
<td></td>
<td>L</td>
<td>Satt284</td>
<td>11.4</td>
<td>3.4*</td>
<td>0.00008</td>
<td>0.39</td>
<td>10</td>
</tr>
<tr>
<td></td>
<td></td>
<td>N</td>
<td>Satt549</td>
<td>37.7</td>
<td>2.8</td>
<td>0.0003</td>
<td>0.33</td>
<td>8</td>
</tr>
<tr>
<td>04 Prin</td>
<td></td>
<td>I</td>
<td>Satt330</td>
<td>61.3</td>
<td>3.0*</td>
<td>0.0001</td>
<td>0.41</td>
<td>9</td>
</tr>
<tr>
<td></td>
<td></td>
<td>L</td>
<td>Satt652</td>
<td>0.0</td>
<td>3.4*</td>
<td>0.00008</td>
<td>0.51</td>
<td>10</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Satt284</td>
<td>11.4</td>
<td>6.7*</td>
<td>0.00000</td>
<td>0.60</td>
<td>19</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Satt678</td>
<td>38.6</td>
<td>2.5</td>
<td>0.0007</td>
<td>0.52</td>
<td>8</td>
</tr>
<tr>
<td></td>
<td></td>
<td>N</td>
<td>Satt549</td>
<td>37.7</td>
<td>3.1*</td>
<td>0.0001</td>
<td>0.40</td>
<td>9</td>
</tr>
</tbody>
</table>
## Table 11. (Continued)

<table>
<thead>
<tr>
<th>Trait</th>
<th>Env‡‡</th>
<th>LG††</th>
<th>Marker</th>
<th>Position</th>
<th>LOD</th>
<th>P value</th>
<th>A</th>
<th>R²(%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>SFP§ (days)</td>
<td>03 Lex</td>
<td>B2</td>
<td>Satt304</td>
<td>32.5</td>
<td>2.6</td>
<td>0.0005</td>
<td>-1.27</td>
<td>8</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>F</td>
<td>Satt335</td>
<td>72.6</td>
<td>3.9*</td>
<td>0.00002</td>
<td>1.41</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>N</td>
<td>Satt584</td>
<td>7.1</td>
<td>2.4</td>
<td>0.0008</td>
<td>-1.17</td>
</tr>
<tr>
<td></td>
<td>04 Lex</td>
<td>F</td>
<td>Satt335</td>
<td>72.6</td>
<td>1.3</td>
<td>0.01</td>
<td>0.77</td>
<td>4</td>
</tr>
<tr>
<td>MAT# (days)</td>
<td>03 Lex</td>
<td>A1</td>
<td>Satt364</td>
<td>3.5</td>
<td>2.5</td>
<td>0.0007</td>
<td>-1.89</td>
<td>7</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>F</td>
<td>Satt335</td>
<td>72.6</td>
<td>1.4</td>
<td>0.01</td>
<td>0.88</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>N</td>
<td>Satt584</td>
<td>7.1</td>
<td>2.1</td>
<td>0.002</td>
<td>-1.13</td>
</tr>
<tr>
<td></td>
<td>04 Lex</td>
<td>F</td>
<td>Satt335</td>
<td>72.6</td>
<td>4.0*</td>
<td>0.00002</td>
<td>1.82</td>
<td>12</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>N</td>
<td>Satt584</td>
<td>7.1</td>
<td>3.1*</td>
<td>0.0002</td>
<td>-1.67</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>O</td>
<td>Satt331</td>
<td>49.7</td>
<td>2.7</td>
<td>0.0004</td>
<td>1.93</td>
</tr>
<tr>
<td>HT† (cm)</td>
<td>04 Lex</td>
<td>F</td>
<td>Satt144</td>
<td>91.9</td>
<td>3.1*</td>
<td>0.0002</td>
<td>8.11</td>
<td>10</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>L</td>
<td>Satt652</td>
<td>0</td>
<td>2.7</td>
<td>0.0005</td>
<td>5.60</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Satt284</td>
<td>11.4</td>
<td>2.1</td>
<td>0.002</td>
<td>4.33</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>N</td>
<td>Satt549</td>
<td>37.7</td>
<td>2.1</td>
<td>0.002</td>
<td>3.95</td>
</tr>
<tr>
<td></td>
<td>04 Prin</td>
<td>B2</td>
<td>Satt318</td>
<td>35.4</td>
<td>2.0</td>
<td>0.02</td>
<td>3.19</td>
<td>6</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>F</td>
<td>Satt595</td>
<td>48.9</td>
<td>2.4</td>
<td>0.0009</td>
<td>3.36</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>G</td>
<td>Satt144</td>
<td>91.9</td>
<td>3.2*</td>
<td>0.0001</td>
<td>6.59</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>L</td>
<td>Satt284</td>
<td>11.4</td>
<td>2.7</td>
<td>0.0004</td>
<td>3.95</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>O</td>
<td>Satt420</td>
<td>11.8</td>
<td>2.4</td>
<td>0.0009</td>
<td>-3.43</td>
</tr>
</tbody>
</table>

§ seed yield, kg/ha, was measured within each line
£ 100 seed weight, g adjusted to a 13% seed water content
# days from 30 August to the date when 95% of the pods had reached a final pod color
† height, distance from the soil to the tip of the main stem at maturity
‡ lodging, 1= all plants standing erect, 5= all plants prostrate
§ seed filling period, number of days from the date of R5 to the date of R7
‡‡ each location-year combination was considered an environment
Lex- Lexington, Prin- Princeton, 03- 2003, 04- 2004
### MR estimated additive effect, the additive effect (sign and magnitude) was that calculated for the substitution of a single *G. soja* allele for a *G. max* allele at the indicated QTL peak position
$$ Logarithm of odds score, the LOD score significance criterion was 3 for MR analysis
†† Linkage group
Table 12. Putative QTL identified by the simple interval mapping (SIM) method for the single environments. The phenotypic data were the means of the 147 BC$_2$F$_4$ in population A that segregated about 7.86MM:1SS for each of 121 markers. The mean map position of the significant QTL was established by the map position of the peak LOD score exceeding 3 in the interval between the two indicated flanking markers. A one-LOD fall-off (from the QTL peak) method was used to estimate the left and right flanking map positions of a confidence interval surrounding the mean QTL map position.

<table>
<thead>
<tr>
<th>Trait</th>
<th>Env†‡</th>
<th>LG††</th>
<th>Marker On Left</th>
<th>QTL Left flank</th>
<th>QTL Mean position</th>
<th>QTL Right Flank</th>
<th>Marker On Right</th>
<th>LOD$$</th>
<th>A##</th>
<th>R² (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>SY³ (Kg/ha)</td>
<td>04Lex</td>
<td>A1</td>
<td>Satt050</td>
<td>33</td>
<td>42</td>
<td>45</td>
<td>Satt511</td>
<td>2.5</td>
<td>214.40</td>
<td>8</td>
</tr>
<tr>
<td></td>
<td></td>
<td>J</td>
<td>Satt285</td>
<td>0</td>
<td>0</td>
<td>5</td>
<td>Satt414</td>
<td>2.8</td>
<td>-235.2</td>
<td>9</td>
</tr>
<tr>
<td></td>
<td>03Prin</td>
<td>A1</td>
<td>Satt050</td>
<td>29</td>
<td>41</td>
<td>45</td>
<td>Satt511</td>
<td>1.5</td>
<td>158.35</td>
<td>5</td>
</tr>
<tr>
<td></td>
<td></td>
<td>J</td>
<td>Satt529</td>
<td>33</td>
<td>37</td>
<td>41</td>
<td>Satt431</td>
<td>5.5**</td>
<td>372.40</td>
<td>16</td>
</tr>
<tr>
<td>100- SW⁶ (g)</td>
<td>04Lex</td>
<td>B2</td>
<td>Satt577</td>
<td>5</td>
<td>11</td>
<td>19</td>
<td>Satt168</td>
<td>2.5</td>
<td>-1.48</td>
<td>8</td>
</tr>
<tr>
<td></td>
<td></td>
<td>D2</td>
<td>Satt002</td>
<td>3</td>
<td>18</td>
<td>24</td>
<td>Satt582</td>
<td>3.9**</td>
<td>-1.15</td>
<td>11</td>
</tr>
<tr>
<td></td>
<td></td>
<td>G</td>
<td>Satt324</td>
<td>30</td>
<td>35</td>
<td>42</td>
<td>Satt594</td>
<td>3.3*</td>
<td>-1.26</td>
<td>10</td>
</tr>
<tr>
<td></td>
<td></td>
<td>H</td>
<td>Satt541</td>
<td>29</td>
<td>44</td>
<td>50</td>
<td>Satt142</td>
<td>4.3**</td>
<td>-1.46</td>
<td>13</td>
</tr>
<tr>
<td></td>
<td></td>
<td>L</td>
<td>Satt284</td>
<td>6</td>
<td>11</td>
<td>21</td>
<td>Satt481</td>
<td>2.3</td>
<td>-0.81</td>
<td>7</td>
</tr>
<tr>
<td></td>
<td></td>
<td>M</td>
<td>Satt551</td>
<td>42</td>
<td>55</td>
<td>67</td>
<td>Satt346</td>
<td>2.6</td>
<td>1.01</td>
<td>8</td>
</tr>
<tr>
<td></td>
<td></td>
<td>N</td>
<td>Satt549</td>
<td>36</td>
<td>40</td>
<td>45</td>
<td>Satt339</td>
<td>3.5*</td>
<td>-1.06</td>
<td>11</td>
</tr>
<tr>
<td></td>
<td>03Prin</td>
<td>B1</td>
<td>Satt426</td>
<td>0</td>
<td>0</td>
<td>4</td>
<td>Satt251</td>
<td>2.4</td>
<td>-0.82</td>
<td>8</td>
</tr>
<tr>
<td></td>
<td></td>
<td>B2</td>
<td>Satt577</td>
<td>5</td>
<td>12</td>
<td>23</td>
<td>Satt168</td>
<td>2.5</td>
<td>-1.37</td>
<td>8</td>
</tr>
<tr>
<td></td>
<td></td>
<td>G</td>
<td>Satt324</td>
<td>29</td>
<td>34</td>
<td>42</td>
<td>Satt594</td>
<td>3.3*</td>
<td>-1.21</td>
<td>10</td>
</tr>
<tr>
<td></td>
<td></td>
<td>H</td>
<td>Satt541</td>
<td>28</td>
<td>45</td>
<td>51</td>
<td>Satt142</td>
<td>3.9**</td>
<td>-1.40</td>
<td>12</td>
</tr>
<tr>
<td></td>
<td></td>
<td>N</td>
<td>Satt538</td>
<td>36</td>
<td>40</td>
<td>45</td>
<td>Satt339</td>
<td>3.5*</td>
<td>-1.06</td>
<td>11</td>
</tr>
<tr>
<td>LDG⁷ (score)</td>
<td>04Lex</td>
<td>A2</td>
<td>Satt315</td>
<td>9</td>
<td>27</td>
<td>41</td>
<td>Satt508</td>
<td>2.2</td>
<td>0.63</td>
<td>7</td>
</tr>
<tr>
<td></td>
<td></td>
<td>L</td>
<td>Satt652</td>
<td>2</td>
<td>7</td>
<td>21</td>
<td>Satt284</td>
<td>3.9**</td>
<td>0.57</td>
<td>12</td>
</tr>
<tr>
<td></td>
<td></td>
<td>N</td>
<td>Satt387</td>
<td>28</td>
<td>35</td>
<td>41</td>
<td>Satt549</td>
<td>2.6</td>
<td>0.41</td>
<td>8</td>
</tr>
<tr>
<td></td>
<td>03Prin</td>
<td>G</td>
<td>Satt594</td>
<td>31</td>
<td>42</td>
<td>47</td>
<td>Satt199</td>
<td>3.7**</td>
<td>0.54</td>
<td>11</td>
</tr>
<tr>
<td></td>
<td></td>
<td>I</td>
<td>Satt270</td>
<td>52</td>
<td>60</td>
<td>67</td>
<td>Satt330</td>
<td>2.7</td>
<td>0.38</td>
<td>8</td>
</tr>
<tr>
<td></td>
<td></td>
<td>L</td>
<td>Satt652</td>
<td>3</td>
<td>8</td>
<td>21</td>
<td>Satt284</td>
<td>3.7**</td>
<td>0.48</td>
<td>11</td>
</tr>
<tr>
<td></td>
<td>04Prin</td>
<td>I</td>
<td>Satt270</td>
<td>51</td>
<td>60</td>
<td>67</td>
<td>Satt330</td>
<td>2.5</td>
<td>0.42</td>
<td>8</td>
</tr>
<tr>
<td></td>
<td></td>
<td>L</td>
<td>Satt652</td>
<td>4</td>
<td>8</td>
<td>18</td>
<td>Satt284</td>
<td>7.4**</td>
<td>0.74</td>
<td>21</td>
</tr>
<tr>
<td></td>
<td></td>
<td>N</td>
<td>Satt387</td>
<td>27</td>
<td>35</td>
<td>44</td>
<td>Satt549</td>
<td>2.3</td>
<td>0.38</td>
<td>7</td>
</tr>
</tbody>
</table>
Table 12. (Continued)

<table>
<thead>
<tr>
<th>Trait</th>
<th>Env</th>
<th>LG</th>
<th>Marker On Left</th>
<th>QTL Mean Position</th>
<th>Marker On Right</th>
<th>LOD</th>
<th>A</th>
<th>R² (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>SFP³ (days)</td>
<td>03 Lex</td>
<td>F</td>
<td>Satt335</td>
<td>70</td>
<td>73</td>
<td>76</td>
<td>Satt334</td>
<td>4.5**</td>
</tr>
<tr>
<td></td>
<td>04 Lex</td>
<td>F</td>
<td>Satt335</td>
<td>-</td>
<td>73</td>
<td>-</td>
<td>Satt334</td>
<td>0.8</td>
</tr>
<tr>
<td>MAT⁷ (days)</td>
<td>03 Lex</td>
<td>A1</td>
<td>Satt165</td>
<td>0</td>
<td>2</td>
<td>7</td>
<td>Satt364</td>
<td>2.6</td>
</tr>
<tr>
<td></td>
<td>04 Lex</td>
<td>F</td>
<td>Satt335</td>
<td>66</td>
<td>76</td>
<td>90</td>
<td>Satt334</td>
<td>1.4</td>
</tr>
<tr>
<td></td>
<td>N</td>
<td>Satt584</td>
<td>1</td>
<td>7</td>
<td>15</td>
<td>Satt387</td>
<td>2.9</td>
<td>-1.69</td>
</tr>
<tr>
<td></td>
<td>O</td>
<td>Satt477</td>
<td>42</td>
<td>49</td>
<td>58</td>
<td>Satt331</td>
<td>2.0</td>
<td>1.87</td>
</tr>
<tr>
<td>HT† (cm)</td>
<td>04 Lex</td>
<td>F</td>
<td>Satt334</td>
<td>88</td>
<td>90</td>
<td>-</td>
<td>Satt144</td>
<td>2.6</td>
</tr>
<tr>
<td></td>
<td>L</td>
<td>Satt652</td>
<td>0</td>
<td>5</td>
<td>10</td>
<td>Satt284</td>
<td>3.3*</td>
<td>6.82</td>
</tr>
<tr>
<td></td>
<td>04 Prin</td>
<td>F</td>
<td>Satt334</td>
<td>87</td>
<td>90</td>
<td>-</td>
<td>Satt144</td>
<td>3.1*</td>
</tr>
<tr>
<td></td>
<td>G</td>
<td>Satt324</td>
<td>31</td>
<td>38</td>
<td>43</td>
<td>Satt594</td>
<td>2.6</td>
<td>4.75</td>
</tr>
<tr>
<td></td>
<td>L</td>
<td>Satt652</td>
<td>2</td>
<td>8</td>
<td>16</td>
<td>Satt284</td>
<td>3.0*</td>
<td>4.84</td>
</tr>
<tr>
<td></td>
<td>O</td>
<td>Satt420</td>
<td>3</td>
<td>12</td>
<td>21</td>
<td>Satt478</td>
<td>2.3</td>
<td>-3.44</td>
</tr>
</tbody>
</table>

* *, ** significant effect at α=0.05, or α=0.01, respectively
§ seed yield, kg/ha, was measured within each line
₤ 100 seed weight, g adjusted to a 13% seed water content
# days from 30 August to the date when 95% of the pods had reached a final pod color
† height, distance from the soil to the tip of the main stem at maturity
‡ lodging, 1= all plants standing erect, 5= all plants prostrate
$ seed filling period, number of days from the date of R5 to the date of R7
‡‡ each location-year combination was considered an environment
Lex- Lexington, Prin- Princeton, 03- 2003, 04- 2004
# SIM estimated additive effect, the additive effect (sign and magnitude) was that calculated for the substitution of a single G. soja allele for a G. max allele at the indicated QTL peak position
$$ Logarithm of odds score, the LOD score significance criterion was LOD>3.6 corresponding to P<0.01, and LOD>3.0 corresponding to P<0.05.
†† Linkage group
Table 13. QTL associated with seed yield, seed filling period and maturity in the *G. max × G. soja* soybean population. Allelic means of seed yield, seed filling period and maturity for the mapping and confirmation population across the testing environments.

<table>
<thead>
<tr>
<th>Pop†</th>
<th>Env‡</th>
<th>Traits</th>
<th>Marker</th>
<th>Allelic means</th>
<th>Effect</th>
<th>Increase (%)</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>03 Lex</td>
<td>SFP§ (days)</td>
<td>Satt335</td>
<td>35.3 38.6</td>
<td>3.3</td>
<td>9.3</td>
<td>0.0089</td>
</tr>
<tr>
<td></td>
<td></td>
<td>MAT° (days)</td>
<td>Satt335</td>
<td>29.0 30.7</td>
<td>1.7</td>
<td>5.9</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td></td>
<td>03 Prin</td>
<td>SY$^$ (kg/ha)</td>
<td>Satt511</td>
<td>3451.0 3709.8</td>
<td>258.8</td>
<td>7.5</td>
<td>0.0008</td>
</tr>
<tr>
<td></td>
<td>04 Lex</td>
<td>SY(kg/ha)</td>
<td>Satt511</td>
<td>3527.4 3896.7</td>
<td>369.3</td>
<td>10.5</td>
<td>0.017</td>
</tr>
<tr>
<td></td>
<td></td>
<td>SFP(days)</td>
<td>Satt335</td>
<td>39.2 40.5</td>
<td>1.3</td>
<td>3.3</td>
<td>0.19</td>
</tr>
<tr>
<td></td>
<td></td>
<td>MAT(days)</td>
<td>Satt335</td>
<td>25.0 28.7</td>
<td>3.7</td>
<td>14.8</td>
<td>0.0007</td>
</tr>
<tr>
<td></td>
<td>04 Prin</td>
<td>SY(kg/ha)</td>
<td>Satt511</td>
<td>3457.0 4092.9</td>
<td>635.9</td>
<td>18.4</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>Combined</td>
<td></td>
<td>SY(kg/ha)</td>
<td>Satt511</td>
<td>3478.4 3899.7</td>
<td>421.3</td>
<td>12.1</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td></td>
<td></td>
<td>SFP(days)</td>
<td>Satt335</td>
<td>37.1 39.6</td>
<td>2.5</td>
<td>6.7</td>
<td>0.019</td>
</tr>
<tr>
<td></td>
<td></td>
<td>MAT(days)</td>
<td>Satt335</td>
<td>27.0 29.8</td>
<td>2.8</td>
<td>10.4</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>B</td>
<td>03 Lex</td>
<td>SY(kg/ha)</td>
<td>Satt511</td>
<td>2432.7 2599.5</td>
<td>166.8</td>
<td>6.9</td>
<td>0.088</td>
</tr>
<tr>
<td></td>
<td></td>
<td>SFP(days)</td>
<td>Satt335</td>
<td>35.3 36.0</td>
<td>0.7</td>
<td>2.0</td>
<td>0.27</td>
</tr>
<tr>
<td></td>
<td></td>
<td>MAT(days)</td>
<td>Satt335</td>
<td>29.1 30.3</td>
<td>1.2</td>
<td>4.1</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td></td>
<td>03 Prin</td>
<td>SY(kg/ha)</td>
<td>Satt511</td>
<td>3273.5 3235.0</td>
<td>-38.5</td>
<td>-1.2</td>
<td>0.70</td>
</tr>
<tr>
<td></td>
<td>04 Lex</td>
<td>SY(kg/ha)</td>
<td>Satt511</td>
<td>3555.9 3825.5</td>
<td>269.6</td>
<td>7.6</td>
<td>0.04</td>
</tr>
<tr>
<td></td>
<td></td>
<td>SFP(days)</td>
<td>Satt335</td>
<td>43.4 45.8</td>
<td>2.4</td>
<td>5.5</td>
<td>0.012</td>
</tr>
<tr>
<td></td>
<td></td>
<td>MAT(days)</td>
<td>Satt335</td>
<td>27.1 28.4</td>
<td>1.3</td>
<td>4.8</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td></td>
<td>04 Prin</td>
<td>SY(kg/ha)</td>
<td>Satt511</td>
<td>2952.6 3331.3</td>
<td>378.7</td>
<td>12.8</td>
<td>0.0035</td>
</tr>
<tr>
<td>Combined</td>
<td></td>
<td>SY(kg/ha)</td>
<td>Satt511</td>
<td>3053.6 3247.7</td>
<td>194.1</td>
<td>6.3</td>
<td>0.037</td>
</tr>
<tr>
<td></td>
<td></td>
<td>SFP(days)</td>
<td>Satt335</td>
<td>39.4 40.9</td>
<td>1.5</td>
<td>3.8</td>
<td>0.015</td>
</tr>
<tr>
<td></td>
<td></td>
<td>MAT(days)</td>
<td>Satt335</td>
<td>28.1 29.4</td>
<td>1.3</td>
<td>4.6</td>
<td>&lt;0.0001</td>
</tr>
</tbody>
</table>

* *, ** significant effect at $\alpha=0.05$, or $\alpha=0.01$, respectively for the A population
* *, ** significant effect at $\alpha=0.25$, or $\alpha=0.1$, respectively for the B population
§ seed yield, kg/ha, was measured within each line
# days from 30 August to the date when 95% of the pods had reached a final pod color
$^\$ seed filling period, number of days from the date of R5 to the date of R7
† A: QTL analysis population
B: QTL validation population
‡ each location-year combination was considered an environment
Lex- Lexington, Prin- Princeton, 03- 2003, 04- 2004
Table 14. QTL associated with seed yield in the *G. max × G. soja* population. Allelic means and class difference for combined allelic type across the testing environments

<table>
<thead>
<tr>
<th>Pop†</th>
<th>Satt511 genotype‡</th>
<th>Satt529 genotype</th>
<th>Allelic means (kg/ha)</th>
<th>Class difference</th>
<th>Effect (kg/ha)</th>
<th>P value</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>m</td>
<td>m</td>
<td>3469.3</td>
<td>mm-ms</td>
<td>-37.2</td>
<td>0.69</td>
</tr>
<tr>
<td></td>
<td>m</td>
<td>s</td>
<td>3506.5</td>
<td>mm-sm</td>
<td>-295.9</td>
<td>0.0004**</td>
</tr>
<tr>
<td></td>
<td>s</td>
<td>m</td>
<td>3765.2</td>
<td>mm-ss</td>
<td>-1213.1</td>
<td>&lt;0.0001**</td>
</tr>
<tr>
<td></td>
<td>s</td>
<td>s</td>
<td>4682.4</td>
<td>ms-sm</td>
<td>-258.7</td>
<td>0.09</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>ms-ss</td>
<td>-1175.9</td>
<td>&lt;0.0001**</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>sm-ss</td>
<td>-917.2</td>
<td>&lt;0.0001**</td>
</tr>
<tr>
<td>B</td>
<td>m</td>
<td>m</td>
<td>3055.3</td>
<td>mm-ms</td>
<td>320.93</td>
<td>0.88</td>
</tr>
<tr>
<td></td>
<td>m</td>
<td>s</td>
<td>2734.3</td>
<td>mm-sm</td>
<td>-154.0</td>
<td>0.11*</td>
</tr>
<tr>
<td></td>
<td>s</td>
<td>m</td>
<td>3209.3</td>
<td>mm-ss</td>
<td>-346.6</td>
<td>0.06**</td>
</tr>
<tr>
<td></td>
<td>s</td>
<td>s</td>
<td>3401.9</td>
<td>ms-sm</td>
<td>-475.0</td>
<td>0.35</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>ms-ss</td>
<td>-667.5</td>
<td>0.09**</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>sm-ss</td>
<td>-192.6</td>
<td>0.20*</td>
</tr>
</tbody>
</table>

*, ** significant effect at $\alpha=0.05$, or $\alpha=0.01$, respectively for the A population
*, ** significant effect at $\alpha=0.25$, or $\alpha=0.1$, respectively for the B population
†A: QTL analysis population; B: QTL validation population
‡m: *G. max* allele; s: *G. soja* allele
Table 15. Analysis of variance of yield, SFP and maturity for the QTL validation population at each environment

<table>
<thead>
<tr>
<th>Environment</th>
<th>Source of variation</th>
<th>SY$§</th>
</tr>
</thead>
<tbody>
<tr>
<td>03Lex</td>
<td>Set††</td>
<td>0.13</td>
</tr>
<tr>
<td></td>
<td>QTL1†</td>
<td>0.009**</td>
</tr>
<tr>
<td></td>
<td>QTL2‡</td>
<td>0.007**</td>
</tr>
<tr>
<td></td>
<td>QTL1×QTL2</td>
<td>0.006**</td>
</tr>
<tr>
<td></td>
<td>Set×QTL1</td>
<td>0.12</td>
</tr>
<tr>
<td></td>
<td>Set×QTL2</td>
<td>0.28</td>
</tr>
<tr>
<td></td>
<td>Set×QTL1×QTL2</td>
<td>0.36</td>
</tr>
<tr>
<td></td>
<td>Rep</td>
<td>0.15</td>
</tr>
<tr>
<td></td>
<td>Rep×Set</td>
<td>&lt;0.0001**</td>
</tr>
<tr>
<td></td>
<td>Line(set×QTL1×QTL2)</td>
<td>0.0008**</td>
</tr>
<tr>
<td>03Prin</td>
<td>Set</td>
<td>0.49</td>
</tr>
<tr>
<td></td>
<td>QTL1</td>
<td>0.84</td>
</tr>
<tr>
<td></td>
<td>QTL2</td>
<td>0.22*</td>
</tr>
<tr>
<td></td>
<td>QTL1×QTL2</td>
<td>0.31</td>
</tr>
<tr>
<td></td>
<td>Set×QTL1</td>
<td>0.04*</td>
</tr>
<tr>
<td></td>
<td>Set×QTL2</td>
<td>0.96</td>
</tr>
<tr>
<td></td>
<td>Set×QTL1×QTL2</td>
<td>0.30</td>
</tr>
<tr>
<td></td>
<td>Rep</td>
<td>0.66</td>
</tr>
<tr>
<td></td>
<td>Rep×Set</td>
<td>0.04*</td>
</tr>
<tr>
<td></td>
<td>Line(set×QTL1×QTL2)</td>
<td>&lt;0.0001**</td>
</tr>
<tr>
<td>04Lex</td>
<td>Set</td>
<td>0.69</td>
</tr>
<tr>
<td></td>
<td>QTL1</td>
<td>0.18*</td>
</tr>
<tr>
<td></td>
<td>QTL2</td>
<td>0.90</td>
</tr>
<tr>
<td></td>
<td>QTL1×QTL2</td>
<td>0.54</td>
</tr>
<tr>
<td></td>
<td>Set×QTL1</td>
<td>0.61</td>
</tr>
<tr>
<td></td>
<td>Set×QTL2</td>
<td>0.77</td>
</tr>
<tr>
<td></td>
<td>Set×QTL1×QTL2</td>
<td>0.44</td>
</tr>
<tr>
<td></td>
<td>Rep</td>
<td>0.03*</td>
</tr>
<tr>
<td></td>
<td>Rep×Set</td>
<td>0.11</td>
</tr>
<tr>
<td></td>
<td>Line(set×QTL1×QTL2)</td>
<td>&lt;0.0001**</td>
</tr>
<tr>
<td>04Prin</td>
<td>Set</td>
<td>0.86</td>
</tr>
<tr>
<td></td>
<td>QTL1</td>
<td>0.10*</td>
</tr>
<tr>
<td></td>
<td>QTL2</td>
<td>0.34</td>
</tr>
<tr>
<td></td>
<td>QTL1×QTL2</td>
<td>0.67</td>
</tr>
<tr>
<td></td>
<td>Set×QTL1</td>
<td>0.12</td>
</tr>
<tr>
<td></td>
<td>Set×QTL2</td>
<td>0.67</td>
</tr>
<tr>
<td></td>
<td>Set×QTL1×QTL2</td>
<td>0.24</td>
</tr>
<tr>
<td></td>
<td>Rep</td>
<td>0.03*</td>
</tr>
<tr>
<td></td>
<td>Rep×Set</td>
<td>0.64</td>
</tr>
<tr>
<td></td>
<td>Line(set×QTL1×QTL2)</td>
<td>&lt;0.0001**</td>
</tr>
<tr>
<td>Env</td>
<td>Source of variation</td>
<td>SFP$</td>
</tr>
<tr>
<td>-----</td>
<td>---------------------</td>
<td>------</td>
</tr>
<tr>
<td>03Lex</td>
<td>Set</td>
<td>0.22</td>
</tr>
<tr>
<td></td>
<td>QTL##</td>
<td>&lt;0.0001**</td>
</tr>
<tr>
<td></td>
<td>Set × QTL</td>
<td>0.36</td>
</tr>
<tr>
<td></td>
<td>Rep</td>
<td>0.24</td>
</tr>
<tr>
<td></td>
<td>Rep × Set</td>
<td>0.08</td>
</tr>
<tr>
<td></td>
<td>Line(set×QTL)</td>
<td>&lt;0.0001**</td>
</tr>
<tr>
<td>04Lex</td>
<td>Set</td>
<td>0.53</td>
</tr>
<tr>
<td></td>
<td>QTL</td>
<td>&lt;0.0001**</td>
</tr>
<tr>
<td></td>
<td>Set × QTL</td>
<td>0.49</td>
</tr>
<tr>
<td></td>
<td>Rep</td>
<td>0.07</td>
</tr>
<tr>
<td></td>
<td>Rep × Set</td>
<td>0.21</td>
</tr>
<tr>
<td></td>
<td>Line(set×QTL)</td>
<td>0.0004**</td>
</tr>
</tbody>
</table>

*, ** significant effect at $\alpha$=0.25, or $\alpha$=0.1, respectively for QTL, QTL1, QTL2, and QTL1 × QTL2, and at $\alpha$=0.05, or $\alpha$=0.01 for the remaining sources of variation

$\text{SFP}$, kg/ha, was measured within each line

$\text{MAT}$, days from 30 August to the date when 95% of the pods had reached a final pod color

$\text{SFP}$, seed filling period, number of days from the date of R5 to the date of R7

$\text{MAT}$, seed filling period and maturity QTL marked by Satt335 on linkage group F

$\text{SFP}$, seed yield QTL marked by Satt529 on linkage group J

$\text{MAT}$, BC₂F₄ lines grouped in sets 1, 2, and 3 according to maturity

$\text{SFP}$, yield QTL marked by Satt511 on linkage group A1

$\text{MAT}$, yield QTL marked by Satt529 on linkage group J

$\text{SFP}$, seed filling period and maturity QTL marked by Satt335 on linkage group F
Table 16. Analysis of variance of yield and SFP for the QTL validation population at combined environments

<table>
<thead>
<tr>
<th>Source of variation</th>
<th>SY§</th>
<th>Source of variation</th>
<th>SFP$</th>
<th>MAT##</th>
</tr>
</thead>
<tbody>
<tr>
<td>Environment</td>
<td>0.008**</td>
<td>Environment</td>
<td>0.0006**</td>
<td>0.48</td>
</tr>
<tr>
<td>Set††</td>
<td>0.72</td>
<td>Set</td>
<td>0.39</td>
<td>&lt;0.0001**</td>
</tr>
<tr>
<td>Env×set</td>
<td>0.07</td>
<td>Env×set</td>
<td>0.35</td>
<td>0.01**</td>
</tr>
<tr>
<td>QTL1†</td>
<td>0.08**</td>
<td>QTL##</td>
<td>&lt;0.0001**</td>
<td>&lt;0.0001**</td>
</tr>
<tr>
<td>QTL2‡</td>
<td>0.16*</td>
<td>Env×QTL</td>
<td>0.10</td>
<td>0.19</td>
</tr>
<tr>
<td>QTL1×QTL2</td>
<td>0.18*</td>
<td>Set×QTL</td>
<td>0.68</td>
<td>0.08</td>
</tr>
<tr>
<td>Env×QTL1</td>
<td>0.22</td>
<td>Env×Set×QTL</td>
<td>0.06</td>
<td>0.01**</td>
</tr>
<tr>
<td>Env×QTL2</td>
<td>0.24</td>
<td>Rep(env)</td>
<td>0.04**</td>
<td>0.42</td>
</tr>
<tr>
<td>Env×QTL1×QTL2</td>
<td>0.30</td>
<td>Rep×Set(env)</td>
<td>0.11</td>
<td>0.15</td>
</tr>
<tr>
<td>Set×QTL1</td>
<td>0.11</td>
<td>Line(set×QTL)</td>
<td>0.03*</td>
<td>&lt;0.0001**</td>
</tr>
<tr>
<td>Set×QTL2</td>
<td>0.99</td>
<td>Env×line(set×QTL)</td>
<td>0.0008**</td>
<td>&lt;0.0001**</td>
</tr>
<tr>
<td>Set×QTL1×QTL2</td>
<td>0.45</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Env×Set×QTL1</td>
<td>0.30</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Env×Set×QTL2</td>
<td>0.34</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Env×Set×QTL1×QTL2</td>
<td>0.16</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Rep(env)</td>
<td>0.01**</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Rep×Set(env)</td>
<td>&lt;0.0001**</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Line(set×QTL1×QTL2)</td>
<td>&lt;0.0001**</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Env×line(set×QTL×QTL)</td>
<td>&lt;0.0001**</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*, ** significant effect at $\alpha=0.25$, or $\alpha=0.1$, respectively for QTL, QTL1, QTL2, and QTL1 × QTL2, and at $\alpha=0.05$, or $\alpha=0.01$ for the remaining sources of variation

§ seed yield, kg/ha, was measured within each line

# days from 30 August to the date when 95% of the pods had reached a final pod color

$ seed filling period, number of days from the date of R5 to the date of R7

††BC$_2$F$_4$ lines grouped in sets 1, 2, and 3 according to maturity

‡§yield QTL marked by Satt511 on linkage group A1

♯yield QTL marked by Satt529 on linkage group J

§§seed filling period and maturity QTL marked by Satt335 on linkage group F
Figure 3. The graphical depiction of the LOD score on the indicated linkage group for the indicated trait, using the marker regression (MR, solid black square), the simple interval mapping (SIM, blue line), and the composite interval mapping (CIM, red line) methods of QTL analysis. The LOD score criterion for significance was 3.0. The additive effect, as estimated by MR (solid black square), SIM (blue line) and CIM (red line) of the indicated linkage groups are presented below the LOD scores on the same linkage group. The additive effect is that calculated for the substitution of a single *G. soja* allele for a *G. max* allele at the indicated locus position. Markers on the indicated linkage group are shown by name, at their map positions on the horizontal axis. A- seed yield, B- seed filling period
A.

Trait: Seed Yield  Linkage Group A1

Position in Morgans

Trait: Seed Yield  Linkage Group A1

Position in Morgans
B.

**Trait: SFP  Linkage Group F**

![Graph of LOD Score](image)

**Position in Morgan**

**Trait: SFP  Linkage Group F**

![Graph of Additive Effects](image)

**Position in Morgans**
Despite soybean’s narrow genetic base, the use of exotic germplasm has not been fully exploited in conventional soybean breeding programs largely due to the success of breeders in crossing elite × elite strains to generate superior cultivars (Thorne and Fehr, 1970). The discovery and mapping of QTL associated with beneficial and novel alleles from exotic parents have been reported in soybean: 1) the population of RILs derived from Mineoy and Archer, and Noir 1 and Archer (Orf et al., 1999a,b), 2) the BC$_2$F$_1$ population derived from soybean line HS-1 and G. soja accession PI 407305 (Concibido, et al. 2003), 3) the BC$_2$ population developed from the G. max cultivar IA2008 and the G. soja plant introduction PI 468916 (Wang et al., 2004), 4) the F$_5$ population developed from a maturity group II cultivar BSR 101 and the experimental line LG82-8379 which is derived from two soybean PIs (Kabelka et al., 2004).

Considering the large number of exotic accessions in seed banks and the limited amount of time and resources available, it is very important to decide which accessions to sample in order to maximize the chance of finding useful genes. Tanksley and McCouch (1997) reported that accessions with DNA profiles most distinct from that of modern germplasm are likely to contain the greatest number of novel alleles.

In this study, the G. soja × G. max inbred backcross population was developed using the parents G. soja PI 245331 and the G. max variety 7499. PI 245331 is a maturity group X accession from Taiwan that was identified by Maughan et al. (1996) as being more diverse from G. max than the remaining G. soja group. The selection of PI 245331 as a very divergent genotype should expand the range of allele discovery beyond the soybean genotypes previously used.

Power of the QTL Test

The availability of molecular markers has led to more than 700 published studies mapping QTL (Coors 2001). In plants, more than 900 QTL have been reported from studies that involved relatively large mapping populations (Bernardo 2002). In order to decrease the risk of reporting false positive QTL (Type-I error), stringent criteria for supporting the validity of reported QTL are essential. In this study the power (1-β) of the experiment was increased in four ways: 1) by setting up a stringent experiment-wise threshold of significance, 2) using multi-year
and multi-location experiments with replications for the mapping population, 3) verifying the QTL in a second population, 4) using MR and SIM analyses followed by CIM analysis to identify QTL that may have been masked by the effect of larger QTL (Zeng, 1994).

The P-value that Map-manager QTX calculates for QTL mapping is the probability of a false positive or the probability of obtaining by chance an LOD at least as great as the observed LOD.

For MR analysis, the P-value applies only to one point in the genetic map. The probability of a false positive anywhere in the genome is much greater (Lander and Kruglyak, 1995). The P-value for a single marker analysis must be well below 0.001 to obtain a genome-wide P-value of 0.05. In this study, for MR analysis, I set stringent detection limits (P<0.0001, LOD>3.0).

In this study, I tried to minimize the risk of reporting false positive QTL (probability of Type-I error, \(\alpha\)) by using permutation tests to define an empirical detection threshold at an experiment-wise P-value<0.01 for the combined SIM and CIM QTL analysis, being more concerned with finding false positive QTL (Type I error, \(\alpha\)) than I were in missing a QTL (Type II error, \(\beta\)). A less stringent experiment-wise significance threshold of P<0.05 was used for the single environment SIM QTL analysis to support the combined data.

Bernardo (2004) conducted research on a F\(_2\) population to address the question: what proportions of declared QTL in plants are false? A simulation study was conducted to determine the false discovery rate (FDR) in an F\(_2\) mapping population by using different significance levels (\(\alpha\)) and heritability estimates. The FDR increased as the significance level increased. The FDR increased to 0.82 when \(\alpha\) was 0.05, heritability was low, and only one QTL controlled the trait. But the increase in heritability led to a decrease in the FDR and an increase in power. As shown in Table 17, the heritability of the traits was high in this study. Setting the experiment-wise threshold of P<0.0001 for MR and P<0.01 for SIM and CIM analysis provides a stringent level of security against Type-I errors.

QTL verification is an important component of QTL introgression and use. For example, Boerma and Mian (1999) were able to confirm only four of eight QTL for seed quality traits when they selected for marker alleles in an independent population of the discovery cross. They advised conducting confirmation experiments prior to initiating breeding programs based on unconfirmed QTL. The 148 lines in population B were genotyped for the SSR loci Satt511, 79
Satt529, and Satt335 near the favorable G. soja QTL yield or SFP alleles. The B population was divided into classes based on the G. soja and G. max alleles contained in the lines to verify that the QTL identified in population A are true and functional in population B.

In order to decrease the risk of reporting false positive QTL, stringent criteria (P<0.01) were used in the QTL discovery population. After a positive QTL was identified in population A, it was verified in population B by using a less stringent criterion (P<0.1), because instead of identifying an unknown QTL in the whole soybean genome, the existing QTL in population A were tested for functionality in population B.

**Mapping Population**

A limited population size in QTL detection experiments may lead to underestimation of QTL number (Beavis, 1994). Many of the advanced backcross introgression populations were 70% larger than this 147 line BC2F4 population (Concibido et al., 2003; Chaky, 2004). The original goal was to produce a 300 line population. But this was not successful due to the difficulties of backcrossing. QTL discovery in advanced backcross populations has both negative and positive features (Kaeppler, 1997). Negatively, the recurrent parent marker class mean becomes more precisely estimated as the number of backcrosses increases, but the donor parent marker class mean becomes less precisely estimated due to the increasing unbalance of the number of members in each group. It is difficult to produce the large enough quantity of F1 seed and an adequately sized mapping population needed in the advanced backcross method. Positively, increased similarity within the population from backcrossing may reduce error variance and improve QTL detection. Assessing the wild species alleles in the elite parent genetic background reduces the impact of epistasis among wild species alleles and improves individual QTL characterization in an elite genetic background.

Wang et al. (2004) developed a series of backcross populations to avoid the disadvantages of the advanced backcross population. The mapping populations consisted of five BC2 populations developed using the G. max variety as the recurrent parent and the G. soja plant introduction as the donor parent. The BC2F4-derived lines from the five populations were used as the mapping population. There were between 57 and 112 BC2F4-derived lines in each population and a total of 468 lines for the five populations. In addition to reducing the number of F1 seed and backcross numbers needed, there are other advantages to using the series backcross method compared to the advanced backcross method. One advantage is a reduction in marker
data collection. After an initial marker screen of the parents and a separate bulk of lines from each population, the populations are tested with only those markers that segregate in the population. With the advanced backcross method, the entire population needs to be tested with all polymorphic markers. But the disadvantage of the series backcross method is the inability to assay the entire *G. soja* genome for QTL because many regions will not segregate in any population. By using the series backcross population, Wang et al. (2004) identified four yield QTL across environments. For these four yield QTL, the *G. max* marker allele was associated with significantly greater yield than the marker allele from *G. soja*. They were unsuccessful in identifying favorable yield QTL alleles from *G. soja*. It is possible that the *G. soja* parent did have one or more positive QTL, but they were not detected in their study because they were not segregating in any population. This is possible because only 52% of the markers that were polymorphic between the parents segregated in at least one population (Wang et al., 2004).

By using an advanced backcross population I identified two positive yield QTL from *G. soja*, one from the combined environments and one from environment Lex04. In my population, segregation was not skewed for any of the 121 markers (Table 2). By waiting until an advanced generation, it would be less likely to detect QTL with epistatic effects, and associated deleterious effects due to linkage drag are also less likely to be observed. In this study, the two yield QTL marked by Satt511 (QTL1) and Satt529 (QTL2) appear to have an interaction effect in addition to the additive effect. As shown in Table 14, allelic mean differences were not significant for mm-ms and ms-sm in both population A and population B. In both populations line with sm yielded significantly more than mm lines and the allelic mean of ss is significantly greater than mm, ms and sm in both populations. Putting these two alleles together is better than each single allele alone even when the *G. soja* allele at Satt529 does not have a significant effect. QTL1 and QTL2 not only have the additive effect, but the interaction effect also exists between them (QTL1 × QTL2 effect (P=0.18), Table 16).

**Seed Yield QTL**

The major objective in this study was to identify trait-enhancing alleles from the wild soybean relative PI 245331 for yield and yield components. In this study, the PI 245331 alleles were favorable for 3 out of 9 (33.3%) yield and yield component QTL, including loci for seed yield, 100-seed weight, seed filling period, maturity, plant height and plant lodging (Table 8). There are reports of identification of positive alleles for yield from wild accessions in other

Concibido et al. (2003) used SSR and amplified fragment length polymorphism (AFLP) markers and the advanced backcross method of QTL mapping to identify a yield QTL that was associated with increased seed yield from G. soja PI 407305. The QTL mapping was conducted in a BC$_2$ population developed according to the advanced backcross method (Tanksley and Nelson 1996). The yield QTL was located near Satt066 on linkage group B2 of the soybean genetic linkage map. In the BC$_2$ population, the QTL allele was associated with a 12% yield increase across testing environments. In order to test the QTL effect, lines containing the QTL were backcrossed into seven genetic backgrounds. In three out of the seven backgrounds, significant (P<0.05) positive effect on yield for the QTL allele from G. soja were observed.

Kabelka et al. (2004) evaluated two soybean PIs as sources of alleles for the enhancement of seed yield in North American cultivars. They identified 15 QTL for seed yield. At nine of the 15 QTL the PI alleles increased seed yield. Satt363 on linkage group C2 explained 10% of the phenotypic variation and increased seed yield 60 kg ha$^{-1}$ across all the environments. Satt358 on linkage group O explained 14% of the phenotypic variation and increased seed yield 47 kg ha$^{-1}$ across 12 environments. The QTL mapping population was grouped into 3 different sets according to maturity. The other 7 yield QTL with the PI alleles increasing seed yield were detected only within specific maturity sets. The alleles contributing to the increase in yield on linkage groups A1, G, K, D2, and M were from one PI, and alleles on linkage groups B2 and H were from the second PI. Within specific maturity sets, these QTL explained 14 to 38% of the phenotypic variation and increased seed yield 60 to 148 kg ha$^{-1}$. One positive yield QTL (LOD=2.6) from the PI was identified within maturity set one marked by Satt225 (95.2cM) (Table 18). This QTL explained 14% of the phenotypic variation and increased seed yield 60 kg ha$^{-1}$ (2.1%) across 12 environments. This Satt225 marker allele also was associated with an increase in days to maturity and with taller plants. The correlation coefficients between seed yield and these traits were small and seed yield was not significantly different between maturity sets. In this dissertation, I report a positive yield QTL from G. soja, marked by Satt511 on linkage group A1 (94.2cM) with an additive yield effect of 190 – 235 kg ha$^{-1}$ depending on the QTL analysis method, accounting for 12-13% of the phenotypic variance of the combined data. This genetic region around Satt511 was not associated with maturity, and there was no obvious
segregation for plant height or any other traits for this allele in my population. Even though these two QTL (Satt225, Kabelka et al. 2004 and Satt511, this dissertation) have different expression for maturity and height, crossing will be necessary to determine if they segregate separately (are different QTL) or if the same QTL has been identified in both studies.

Another positive yield QTL from G. soja was identified in environment 04Prin, marked by Satt529 on linkage group J with increased seed yield of 374-430 kg ha$^{-1}$ and accounting for 10-12% of the phenotypic variance. Mansur and Orf (1996) developed a F$_7$-derived RIL population from crossing Noir1 and Minsoy. They identified a positive yield QTL from Minsoy marked by A060 on linkage group J near my yield QTL Satt529 (Table 18). Marker A060 was also linked to plant height with the positive height effect coming from Noir 1. The QTL at A060 explained 6.1% of the phenotypic variation of yield and 7.6% of the phenotypic variation of plant height, respectively. Specht et al. (2001) developed a F$_{7:11}$-derived RIL population from the cross of Minsoy and Noir 1. A positive yield QTL was identified from Noir 1 marked by G815 on linkage group J (Table 18). That QTL explained 2% of the phenotypic variation. Orf and Chase (1999a) also discovered a QTL on linkage group J in a Minsoy x Noir 1 RIL population where the Minsoy allele, marked by Satt405 (Table 18), increased seed yield and accounted for 9% of the phenotypic variance. These four yield QTL on linkage group J span the length of the chromosome (positions 12.4, 20.8, 41.9, and 73.7cM). Thus the expectation is that these QTL are different. Once again, however, only a directed crossing program will be able to confirm if these QTL are unique.

Seed Filling Period QTL

Higher yields must come from either more seeds per unit area or from a longer seed filling period (Egli, 1998). One SFP QTL was identified for the combined data (Table 8, 9 and 10). The QTL marked by Satt335 was significant (LOD=4.0). The G. soja allele at Satt335 on LG-F (72.6cM) increased SFP duration. The additive effect of the QTL at Satt335 was 1.13 days (LOD=4.0), with a R$^2$ of 12% over combined environments.

No other seed filling period QTL has been identified so far. Orf and Chase (1999a) did identify a yield/seed weight QTL located near my seed filling period QTL. Their yield/seed weight QTL was marked by RFLP marker L050_14 on linkage group F (71.4cM). The mapping population was a RIL population from the cross between Noir 1 and Archer. The favorable allele source was Archer. This QTL explained 16% of the phenotypic variance.
**Maturity QTL**

Time to flowering and maturity in soybean is controlled genetically by the E series of loci: E1, E2 (Bernard, 1971), E3 (Buzzell, 1971), E4 (Buzzell and Voldeng, 1980), E5 (McBlain and Bernard, 1987), E6 (Bonato and Vello, 1999), and E7 (Cober and Voldeng, 2001a). For most of these loci, late flowering and maturity is dominant or partially dominant and early flowering and maturity is recessive, but for the E6 locus, early flowering is dominant.

In this study, two maturity QTL were observed for the combined data by MR analysis with LOD>3.0 (Table 8). QTL at Satt335 on LG-F has an additive effect of 1.4 days (LOD=3.3) with R²= 10%. QTL at Satt584 on LG-N has an additive effect of -1.4 days (LOD=3.2) with R²= 9%.

These QTL at satt584 and satt335 are not E1, E2, and E3 which mapped to linkage groups C2, O, and L respectively (Cregan et al., 1999). E7 is linked to E1 and T (Cober and Voldeng, 2001a) on linkage group C2; thus, QTL marked by Satt584 and Satt335 are not E7.

**QTL Association**

Many regions contained QTL that were significant for more than one trait (Table 8 and Table 9). For example the marker allele Satt335 on linkage group F was associated with an increase in days to maturity and also was associated with an increase in days of seed filling period. The Pearson correlation coefficient between seed maturity and seed filling period in this experiment is 0.39 (P<0.01) (Table 7). It’s likely that the longer seed filling period is caused by the late maturity. Further genetic dissection of the region containing this QTL would be needed to distinguish between pleiotropy or gene linkage.

Another marker allele Satt652 on linkage group L associated with an increase in lodging scores also was associated with an increase in plant height. The Pearson correlation coefficient between lodging and plant height is 0.55 (P<0.01) (Table 7). It is likely that the high lodging score is caused by the tall plants. More research is needed to dissect whether these multiple trait associations are the result of pleiotropy or genetic linkage.

**Future Perspectives**

In this study, I used the advanced backcross method and developed a BC₂F₄ population as the mapping population. The major objective was to identify trait-enhancing alleles from the wild soybean relative PI 245331 for yield and yield components. In my results, the PI 245331 alleles were favorable for 3 out of 9 (33.3%) yield and yield component QTL, including loci for
seed yield, 100-seed weight, seed filling period, maturity, plant height and plant lodging (Table 8).

Three of four environments showed significant (P<0.1) yield increases with homozygosity for the PI 245331 allele at QTL Satt511 (Table 15). For the combined environments, lines that were homozygous for the PI 245331 allele at the QTL locus Satt511 on LG-A1 had a significant effect (P=0.08) (Table 16). These lines yielded 6.3% more (P=0.037) than lines that were homozygous for the *Glycine max* ‘7499’ allele (Table 13). Two out of four locations showed significant (P<0.25) yield increase with homozygosity for the PI 245331 allele marked by Satt529 on linkage group J (Table 15). For the combined environment, the QTL effect was significant (P<0.25) (Table 16). Results from this study indicate that *G. soja* PIs contain QTL alleles that may be of benefit to improving North American soybean cultivars. Of particular interest is the yield QTL marked by Satt511 on linkage group A1. The QTL was significant (LOD=4.3) in the combined analysis across the environments. For this QTL, the PI 245331 allele enhanced yield 190 – 235 kg ha⁻¹ depending on the QTL analysis method. To date, no other studies have detected a yield QTL at this chromosomal location, although it is very close to a QTL identified by Kabelka et al. (2004).

Examples have been reported that QTL identification and subsequent assessment of marker assisted selection occurred successfully in different genetic backgrounds and similar environments (Concibido et al. 2003, Sebolt et al. 2000). But the value of these yield QTL in a broad array of genetic backgrounds and environments is not known. Reyna and Sneller (2001) tried to assess the value of three QTL from northern cultivar Archer for increased yield in southern environments and genetic backgrounds. But, following backcrossing none of the marker effects was significant for any of the three QTL for any trait. Finally, the authors drew the conclusion that it may be difficult to capture the value assigned to QTL alleles derived from diverse parents with variable genetic value when the alleles are introgressed into populations with different genetic backgrounds and when tested in different environments.

It is also important to show that the *G. soja* alleles are unique and not present in the *G. max* germplasm. Sebolt et al. (2000) tried to evaluate the effect of one *G. soja* protein QTL in three genetic backgrounds. The test populations were developed by crossing a line from the initial mapping population with the cultivars Parker and Kenwood and the experimental high protein line C1914. The cultivars are high yielding and are representative of cultivars in
commercial production. The protein QTL allele from *G. soja* was associated with an increase in protein concentration in the Parker and Kenwood populations but not in the C1914 population. This suggested that there may be a gene conferring increased protein concentration in C1914 but not in Parker or Kenwood that is allelic with the *G. soja* gene.

In order to confirm the yield QTL identified in this study in other genetic backgrounds, lines containing these QTL will be crossed to four soybean lines that are high yielding and of diverse parentage. The four lines will be selected to represent the cultivars in commercial production and to insure confirmation in the best current genetic material. This will also put the confirmed alleles into a useful germplasm source. These lines will also be crossed with the *G. soja* yield allele containing HS1 line deposited in the American Type Collection (Concibido et al., 2003) to see if the alleles are additive. Lines from population A that contain both favorable *G. soja* yield QTL (Satt511, Satt529) are 108A and 130A. Their yields were 4878 and 5007 kg/ha, respectively. However, both lines showed a lodging score greater than 3, and they both exhibited up to 50% pod shattering. These two lines will not be used. Lines from population A that contain either favorable *G. soja* yield QTL and show good agronomic traits will be used for future research. Three lines have been selected. Lines 99A and 102A both contain *G. soja* yield QTL Satt511, and they are the top two yielding lines that contain the allele Satt511. Line 153A contains *G. soja* yield QTL Satt529, and it is the top yielding line containing the allele at Satt529. The other agronomic traits of these three lines are superior to lines 108A and 130A. These three lines will be used to make crosses to elite soybean lines.

More experiments will be needed to test these yield QTL in a broader array of genetic backgrounds and in different environments. Also, closely linked markers or flanking markers to the QTL are needed in future research, because recombination between the markers and the QTL could occur. If recombination occurs between a marker and a QTL, then selection based on the marker will not be effective. Nearby markers are available for the yield QTL on linkage group A1. Polymorphism and linkage will be tested in future experiments. At this time favorable alleles have been identified from *G. soja*. The availability of more SSR markers around the QTL region makes it a good candidate for marker assisted breeding in the future.
Table 17. Mean squares and broad-sense heritability ($H^2$) for the BC$_2$F$_4$ lines of the *G. max × G. soja* mapping population grouped into sets and evaluated in 3 environments

<table>
<thead>
<tr>
<th>Source of variation</th>
<th>SY$^\S$ (Kg/ha)</th>
<th>100-SW$^\S$ (g)</th>
<th>HT$^\T$ (cm)</th>
<th>LDG$^\P$ (score)</th>
<th>MAT$^#$ (days)</th>
<th>SFP$^$ (days)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Environment</td>
<td>148.7</td>
<td>105.1</td>
<td>8431.2</td>
<td>41.2</td>
<td>2095.4</td>
<td>1969.0</td>
</tr>
<tr>
<td>Set$^{††}$</td>
<td>838.3</td>
<td>60.9</td>
<td>1931.3</td>
<td>11.1</td>
<td>1741</td>
<td>132.8</td>
</tr>
<tr>
<td>Set×Env</td>
<td>511.4</td>
<td>2.1</td>
<td>67.0</td>
<td>0.96</td>
<td>60.3</td>
<td>151.2</td>
</tr>
<tr>
<td>Rep(env)</td>
<td>551.5</td>
<td>8.8</td>
<td>255.2</td>
<td>4.2</td>
<td>730.5</td>
<td>256.1</td>
</tr>
<tr>
<td>Set×rep(env)</td>
<td>289.4</td>
<td>1.5</td>
<td>325.2</td>
<td>4.5</td>
<td>12.5</td>
<td>14.1</td>
</tr>
<tr>
<td>Line(set)$^{††}$</td>
<td>194.4</td>
<td>26.6</td>
<td>392.2</td>
<td>3.8</td>
<td>17.0</td>
<td>21.2</td>
</tr>
<tr>
<td>Env×line(set)</td>
<td>54.1</td>
<td>0.83</td>
<td>81.6</td>
<td>0.33</td>
<td>7.9</td>
<td>9.4</td>
</tr>
<tr>
<td>Residual</td>
<td>49.5</td>
<td>0.30</td>
<td>63.2</td>
<td>0.24</td>
<td>4.6</td>
<td>6.3</td>
</tr>
<tr>
<td>$H^2$</td>
<td>0.72</td>
<td>0.97</td>
<td>0.79</td>
<td>0.91</td>
<td>0.53</td>
<td>0.56</td>
</tr>
</tbody>
</table>

$^\S$ seed yield, kg/ha, was measured within each line

$^\S$ 100 seed weight, g adjusted to a 13% seed water content

$^\#$ days from 30 August to the date when 95% of the pods had reached a final pod color

$^\T$ height, distance from the soil to the tip of the main stem at maturity

$^\P$ lodging, 1 = all plants standing erect, 5 = all plants prostrate

$^\$ seed filling period, number of days from the date of R5 to the date of R7

$^{††}$ BC$_2$F$_4$ lines grouped in sets 1, 2, and 3 according to maturity

$^{††}$ BC$_2$F$_4$ derived lines

$^\¥$ Broad-sense heritability calculated by the variance component method on a line-mean basis according to Fehr (1987)
Table 18. Yield QTL identified from unimproved accessions in this and previous reports and their linkage group locations.

<table>
<thead>
<tr>
<th>Reference</th>
<th>Species</th>
<th>Source</th>
<th>Marker</th>
<th>Linkage Group</th>
<th>Location (cM)</th>
<th>LOD</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Concibido (2003)</td>
<td>G. soja</td>
<td>PI407305</td>
<td>Satt066</td>
<td>B2</td>
<td>78.8</td>
<td>2.85</td>
<td>0.0001</td>
</tr>
<tr>
<td>Kabelka (2004)</td>
<td>G. max</td>
<td>LG82-8379 (PI68508 × FC04007B)</td>
<td>Satt363</td>
<td>C2</td>
<td>98.1</td>
<td>--</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td></td>
<td></td>
<td>LG82-8379</td>
<td>Satt358</td>
<td>O</td>
<td>5.4</td>
<td>3.0</td>
<td>0.0003</td>
</tr>
<tr>
<td></td>
<td></td>
<td>LG82-8379</td>
<td>Satt225</td>
<td>A1</td>
<td>95.2</td>
<td>2.6</td>
<td>0.005</td>
</tr>
<tr>
<td></td>
<td></td>
<td>LG82-8379</td>
<td>Satt394</td>
<td>G</td>
<td>43.4</td>
<td>6.0</td>
<td>0.0003</td>
</tr>
<tr>
<td></td>
<td></td>
<td>LG82-8379</td>
<td>Satt544</td>
<td>K</td>
<td>43.3</td>
<td>5.9</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td></td>
<td></td>
<td>LG82-8379</td>
<td>Satt168</td>
<td>B2</td>
<td>55.2</td>
<td>6.7</td>
<td>0.003</td>
</tr>
<tr>
<td></td>
<td></td>
<td>LG82-8379</td>
<td>Satt186</td>
<td>D2</td>
<td>105.4</td>
<td>3.2</td>
<td>0.027</td>
</tr>
<tr>
<td></td>
<td></td>
<td>LG82-8379</td>
<td>Satt142</td>
<td>H</td>
<td>86.5</td>
<td>2.7</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td></td>
<td></td>
<td>LG82-8379</td>
<td>Satt308</td>
<td>M</td>
<td>130.8</td>
<td>--</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>Mansur (1996)</td>
<td>G. max</td>
<td>Minsoy</td>
<td>A060</td>
<td>J</td>
<td>20.8</td>
<td>--</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Specht (2001)</td>
<td>G. max</td>
<td>Noir1</td>
<td>G815</td>
<td>J</td>
<td>73.7</td>
<td>2.7</td>
<td>--</td>
</tr>
<tr>
<td>Orf (1999)</td>
<td>G. max</td>
<td>Minsoy</td>
<td>Satt405</td>
<td>J</td>
<td>12.4</td>
<td>3.8</td>
<td>--</td>
</tr>
<tr>
<td>Li</td>
<td>G. soja</td>
<td>PI245331</td>
<td>Satt511</td>
<td>A1</td>
<td>94.2</td>
<td>4.3</td>
<td>0.00001</td>
</tr>
<tr>
<td></td>
<td></td>
<td>PI245331</td>
<td>Satt529</td>
<td>J</td>
<td>41.9</td>
<td>3.4</td>
<td>0.00008</td>
</tr>
</tbody>
</table>
References


Lark, K.G., and K. Chase. 1995. Interactions between quantitative trait loci in soybean in which trait variation at one locus is conditional upon a specific allele at another. PNAS, 92:4656-4660.


Vita

Dandan Li

Date of Birth: September 26\textsuperscript{th}, 1976
In Shandong, China

Degree awarded
Bachelor of Science in Crop Science
Shandong Agricultural University
Taian, China July 1998
Master of Science in Plant Breeding
Shandong Agricultural University
Taian, China July 2001

Dandan Li