To the Graduate Council:

I am submitting herewith a thesis written by Catherine Nyaguthii Nyinyi entitled "Detection and Validation of Agronomic and Seed Quality Quantitative Trait Loci in Soybean." I have examined the final electronic copy of this thesis for form and content and recommend that it be accepted in partial fulfillment of the requirements for the degree of Master of Science, with a major in Plant Sciences.

Vincent R. Pantalone
Major Professor

We have read this thesis
and recommend its acceptance:

Fred L. Allen

Dean A. Kopsell

Carl E. Sams

Accepted for the Council:

Carolyn R. Hodges
Vice Provost and
Dean of the Graduate School

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

Detection and Validation of Agronomic and Seed Quality Quantitative Trait Loci in Soybean

A Thesis Presented for the Master of Science Degree
The University of Tennessee, Knoxville

Catherine Nyaguthii Nyinyi
August 2008

Acknowledgements

I would like to express my gratitude to all those who have helped me throughout my studies, in conducting research and in the preparation of this thesis.

My major Professor Dr. Vincent Pantalone, thank you for your kindness, encouragement, guidance and generosity of spirit. I hope that we shall continue to work together on more exciting and enlightening projects.

I would also like to thank my committee members, Dr. Fred Allen, Dr. Dean Kopsell and Dr. Carl Sams. Thank you for your continued guidance and support throughout the preparation of this thesis.

Members of the Soybean Team, Benjamin Fallen, Deborah Landau-Ellis, Elizabeth Meyer, Sarita Panthee, Amanda Pierce and David Walker, thank you for providing me with much needed help, suggestions and friendship.

I would also like to thank Dr. David Hyten, Dr. Dilip Panthee, Dr. Arnold Saxton and Allison Stewart for their help and suggestions on various data analyses in the preparation of this thesis.

I would also like to thank my family and friends for their unwavering support and patience throughout my studies.

Support from the United Soybean Board, The Tennessee Soybean Promotion Board and the University of Tennessee Institute of Agriculture is greatly appreciated.

Abstract

Soybean seed quality and agronomic traits are important commercially. Agronomic traits such as yield, plant height, lodging, and adapted maturity have been the primary focus of breeders for many years. Seed quality traits are also important as they affect the market price of soybean. Higher protein soybean historically is valued more per unit. It is the goal of plant breeders therefore to simultaneously improve seed quality and agronomic traits. Seed quality and agronomic traits are quantitative traits whose inheritance is governed by many genes, and whose expression is subject to environmental variation. Furthermore, negative correlations between yield and protein, and protein and oil make it even more difficult to select for these traits. Molecular breeding tools such as quantitative trait loci (QTL) can provide breeders with a more direct method of selection for traits at the molecular level. QTL can however be misleading as they are subject to type I and type II errors. QTL validation studies are essential to marker assisted programs as they negate the need for individual breeders to validate every QTL of interest. The purpose of this study was to validate previously reported seed quality and agronomic trait QTL in an independent population derived from an Essex x Williams 82 cross. We were able to validate QTL for all traits and detected novel QTL that may be useful to breeders.

Table of contents

List of Tables

TABLE PAGE

Part 1

Introduction and Literature Review

Introduction

Soybean [*Glycine max* (L.) Merrill] is an annual legume originally domesticated in China (ca.1700-1100 B.C.) and is now grown worldwide (Hymowitz, 1987). The world's largest producers of soybean are: the United States of America (USA), Brazil, Argentina, China and India (Wilcox, 2004). Soybean was introduced in the USA as a forage crop, but is now primarily grown as a grain crop (Smith and Huyser, 1987). USA soybean farm cash receipts for the year 2006 were valued at $20.4 billion; clearly establishing soybean production as a highly valuable enterprise (http://www.ers.usda.gov/News/soybeancoverage.htm verified 13 April 2008).

Soybean seed contains about 40% protein and 20% oil, both of which are important commercially. In addition, soybeans also contain phytochemicals (e.g., phytoestrogens) which have received much attention due to claims of numerous health benefits (Liu, 1997).

Soybeans may be divided into "food beans," "commodity beans" and a third vegetable type soybean (edamame). These differ in seed quality characteristics that are selected based on the end product desired. Food grade beans are grown for direct human consumption or for processing into food products. They typically have a lighter seed coat color and clear hilum, high protein and low oil content and are generally lower yielding than oil beans. There are also specifications for seed size depending upon the product intended; for example producers of natto (a fermented traditional Japanese food) prefer round, small-seeded soybeans with soft texture (Liu, 1997).

Most traditional soy foods originated in Asia. Natto, tempeh, tofu, and soymilk are common examples. Contemporary soy-foods include: protein shakes, meat analogs, breakfast foods, soy coffee and dietary supplements (http://www.soyfoods.com/soyfoodsdescriptions/descriptions.html verified 14 April 2008). Soybean and soy product consumption has increased due to health benefit claims attributed to soy proteins and isoflavones including but not limited to: lowering low density lipoprotein (LDL) cholesterol levels, preventing breast and prostate cancers, diabetes, obesity and osteoporosis (Friedman and Brandon 2001; Anderson et al., 1999). Isoflavones have also received much attention due to claims that they may have detrimental effects when consumed in high doses. Reports of abnormalities in animals and humans have been made, but remain mainly anecdotal until more research is conducted to establish threshold values of the harmful effects, if any, of isoflavones.

Commodity beans are grown for soybean meal (SBM) production, a process that produces soybean oil as a by-product. SBM is a major component of livestock feed. In 2007, SBM accounted for 69% of world protein meal while soybean oil accounted for 30% of the world's vegetable oil consumption (Soystats, 2008). Soybean oil has both dietary and industrial applications. Soybean oil can be processed into margarine, soy vegetable oil, mayonnaise, biodiesel, lubricants and other edible or industrial oil products.

Soybean oil primarily contains the following fatty acids: palmitic acid (16:0), stearic acid (18:0), oleic acid (18:1), linoleic acid (18:2), and linolenic acid (18:3). The first number in the parentheses relates to the number of carbon atoms and the second

number to the number of C=C bonds. The concentration and characteristics of each fatty acid determine the quality of products produced. For example linolenic acid is prone to oxidation which turns the oil product rancid reducing its shelf life. Partial hydrogenation of soybean oil is used to reduce the levels of linolenic acid increasing the shelf life of the product (Lusas, 2004; Pantalone et al., 2004). Hydrogenation of soybean oil, however, results in the formation of trans-fatty acids which are coronary disease risk factors. Through plant breeding, low linolenic acid soybean varieties have been developed to counter this problem. Vistive™ soybeans (Monsanto Company) contain <3% linolenic acid and will be used for the production of soybean oils which contain less trans fats (http://www.vistive.com/about.aspx verified 13 April. 2008).

The value of soybean is attributed to its many qualities and usefulness. To increase soybean value, plant breeders focus on the improvement of those qualities to satisfy consumer preferences. Plant breeders have achieved much success in improving both agronomic and seed quality traits. Conventional plant breeders select for superior genotypes based on phenotypic traits. Markers are distinguishing features between individuals used to differentiate them. They may be simple characters such as flower color or pubescence color. Such characters are called qualitative traits because a single gene, or a few genes govern their expression and they often follow simple Mendelian inheritance (Bernardo, 2002). In contrast, most economically important traits (such as seed yield, protein, and oil concentration) are quantitative traits governed by many genes, each with small or large additive effects.

Conventional breeders use various selection methods for crop improvement. Recurrent selection, for example, is a method employed to increase the number of

favorable alleles in a given population enabling subsequent selection of lines or individuals containing the best combination of alleles. Using recurrent selection, Wilcox (1998) successfully selected for higher seed protein content and increased seed protein from 438 to 484 g kg^{-1} in a population developed from two F_2 populations from the crosses 'Beeson' (ms2ms2) x 'Corsoy' and 'Wells' (ms2ms2) x 'Hark' that were blended with a high protein selection from the cross 'Cutler 71' x 'Pando'. After cycle six (C6) there was no gain in protein content which suggested that the population had no further genetic variability to increase protein. Conventional breeding though, successful in the improvement of crops, is slow and time consuming.

Molecular breeding tools have improved the accuracy and reduced the time required for trait selection. Breeders have the added advantage of being able to recognize individual genes, or chromosomal regions that affect expression and inheritance of a given trait. A quantitative trait such as yield, height or seed size is controlled by many genes and "[…] *do not exhibit classic Mendelian inheritance* (Beavis 1998, p.145)." This makes it difficult to map or even select for individual genes affecting inheritance and expression. There can be greater breeding success when chromosomal regions identified with the trait of interest, referred to as quantitative trait loci (QTL) are mapped. Molecular markers are proteins or DNA that can be useful in detecting polymorphisms that can distingish between individuals (Prince and Ogundiwin, 2004). Molecular DNA markers are differences in genetic sequences that can be used to differentiate between individuals. Examples of molecular DNA markers are: Restriction Fragment Length Polymorphisms (RFLPs), Amplified Fragment Length Polymorphisms (AFLPs), Simple Sequence Repeats (SSRs) and Single Nucleotide Polymorphisms (SNPs). DNA markers

can be used to identify the location of a gene or genes on a chromosome (locus) or to identify QTL.

Literature Review

QTL detection studies are an integral part of marker-assisted selection (MAS) breeding programs. Errors do occur in QTL studies resulting in erroneous association of QTL with a trait or a QTL not being detected at all. These errors may be limited by conducting QTL confirmation studies. Fasoula et al. (2004) point out that due to lack of QTL confirmation studies MAS breeding programs often have to conduct preliminary trials to validate existence of QTL before using reported and unconfirmed QTL in breeding experiments. Conducting confirmation experiments using independent populations and diverse environments to verify the existence of reported QTL can eliminate this initial step (Fasoula et al., 2004). In addition, population sizes should be varied as necessary. According to Beavis (1994), detection of QTL with small genetic effects requires large population sizes while those with large effects may not.

Some QTL confirmation studies have been conducted with varying results, which further reiterates the importance of a verification step to support the use of QTL in breeding programs if true genetic gains are to be realized. Sebolt et al. (2000) conducted a confirmation study of QTL controlling protein from a *G. max* x *G. soja* population previously mapped by Diers et al. (1992). In Diers' study QTL controlling seed protein and oil were mapped on linkage groups (LGs) E and I. A BC_3F_4 line derived from a F_2 (A81-356022 x PI468916) x A81-356022 (recurrent parent) was crossed to two cultivars 'Parker' and 'Kenwood' and the experimental line C1914 to test for the protein QTL. The

BC_3F_4 line was selected because it was homozygous for the *G. soja* region on linkage group (LG) I where the protein QTL mapped. Parker, Kenwood and C1419 were used as representatives of cultivars in commercial production because they are high yielding and have average protein content. In the Parker and Kenwood populations, the protein QTL on LG I from PI468916 was associated with increased protein content and a yield penalty. A similar association was not made in the C1419 population and the authors speculate that it contains a gene allelic to that of *Glycine soja* on LG I but does not contain the yield reducing allele from PI 468916. The QTL previously detected on LG E was not significantly associated with protein in this study (Sebolt et al., 2000).

Fasoula et al. (2004) sought to confirm previously reported QTL for seed oil, protein and weight in an independent population of PI97100 x 'Coker 237' with the same RFLP markers used in the original study. They did the same for previously reported QTL in a 'Young' x PI416937 population using SSR markers mapped in the same region as the original RFLP markers.

For the PI97100 X Coker 237 population, protein QTL identified by RFLP markers E/A454-1 and UNK/A132-4 were confirmed and are designated cqProt-001 and cqProt-002 respectively. However, two other loci were not confirmed because one, (H/A566-2), was not found to be significant and the other, (K/A065-1), was not detected in the confirmation population. Three RFLP markers found in the mapping population were used to validate oil in the confirmation population. RFLP Markers C1/A063-1 and H/A566-2 were confirmed and assigned the names cqOil-001 and cqOil-002 respectively. The third RFLP marker G/L154-2 was not found to be significant in the confirmation population. None of the three seed weight QTL found in the original mapping study were

found in the confirmation population. The LG denoted UNK was unlinked to known linkage groups.

In the Young x PI416937 confirmation population, three QTL for seed protein were tested. None of the SSR loci linked to the RFLP marker that detected the QTL were confirmed. The authors note that the three QTL in the original mapping population were detected in three different environments and were probably subject to sample size differences. One QTL (L/A023-1) for oil named cqOil-003 was confirmed while the other two were not confirmed in the independent population. The investigators confirmed two of three QTL previously found in the mapping population for seed weight. The QTL on linkage groups G (Satt303) and E (Satt263) were named cqSd wt-001 and cqSd wt-002, respectively. The third SSR marker on LG C1 (Satt396) was not detected in the confirmation population. The data presented by Fasoula et al. (2004) indicates the importance of using independent meiotic events, adequate sample sizes and varying environments to test new QTL so as to reduce reported QTL inconsistencies and provide breeders with more reliable information before investing valuable resources for marker-assisted selection.

Panthee et al. (2005) conducted a study on QTL for seed size, seed protein and oil concentration in an N87-984-16 by TN93-99 population. They reported seed quality QTL on LGs D1a, D1b, D2, G, H and O. Their study highlights the occurrence of environmentally sensitive QTL that are only expressed in certain conditions. These QTL may be useful when selecting for a specific environment, but may be difficult to adapt when selecting for diverse environments (Panthee et al., 2005).

In an 'Essex' x 'Williams' RIL population, Hyten et al. (2004a) detected QTL for seed size, oil content and protein content. Six oil concentration QTL detected on linkage groups C2, D1a, D2, L and M. QTL on L and M were significant in all six environments. Seed protein QTL were found on LG C2, D1b, F, K and M. QTL on LG C2, D1b, F and one on K were found to be significant in all six environments while the QTL on LG M was only significant in one. Five QTL associated with seed size were found on LGs D1a, F, G, I, K, and L. QTL on LGs D1a, F, G and L were significant in all six environments while those on I and K were significant in one. Markers used to map QTL in the Essex x Williams population were then used to track the inheritance of the seed quality traits in 'A3127', an Asgrow™ (Monsanto Company) cultivar, developed from an Essex by Williams cross. A3127 inherited some QTL that increased, protein, oil and seed size that did not negatively impact yield (Hyten et al., 2004a). This study shows that QTL mapping may be used to successfully track the inheritance of traits from parents to progeny in a short period of time; a useful tool in a marker assisted breeding program.

Stefaniak et al. (2006) compared recombination in 10 cultivars and 156 RILs all from a Williams x Essex cross. Recombination was scored by listing the markers for each chromosome in the order found on the genetic map then counting crossing over by following each line's marker scores and identifying where an allele changed from one parent to the other. Chi square (χ^2) significance analyzed overall for the RIL set revealed that parental alleles were inherited in the expected 1:1 ratio. However, 8% of the markers had significant χ^2 values ($p < 0.05$) on LGs B2, C2 and L showing deviation from the expected 1:1 ratio. On linkage groups B2 and L, Williams alleles accounted for the significance, while Essex alleles caused significance in LG C2. Twelve markers had

significant χ^2values ($p < 0.01$) of which five on C2 had many Essex alleles. The remaining seven makers, five of which had more Williams alleles, were all on different LGs. Using molecular markers, the authors showed that soybean cultivars arising from Williams x Essex crosses resulted from more recombination events than unselected lines from the same cross. This may illustrate the way alleles are inherited upon selection. This method may be used to track inheritance of traits in populations where phenotypic data already exists to fast track superior progeny in breeding programs.

While mapping regions governing total oil is useful, it is important to recognize that soybean oil contains the following fatty acids: palmitic acid (16:0), stearic acid (18:0), oleic acid (18:1), linoleic acid (18:2) and linolenic acid (18:3) which are found in different concentrations in the seed depending on cultivar and environment. The range of fatty acids is reported as follows: palmitic (8-17%), stearic (3-30%), oleic (25-60%), linoleic (25-60%) and linolenic (2-15%) (Liu, 1997). Differences in fatty acid composition influence the physical properties of fats and oils. Soybean cultivars with different fatty acid profiles have been produced to suit consumer preferences depending on end use.

Hyten et al. (2004b) found QTL for modifier genes of the five fatty acids in an Essex x Williams population. QTL were mapped to LGs C2, D1b, D2, F, K and L. QTL on LG L had effects on 16:0, 18:1, 18:2 and 18:3. New QTL for 18:1 were found on LG D1b and L. Linkage group F showed new QTL for 18:2 and 18:3, and the distance between the two is large enough to consider them to be two separate QTL. The concentration of the different fatty acids in soybean oil influences the characteristics of the products processed from it. The QTL identified in their study are candidate QTL for

confirmation studies. Information on the QTL affecting the expression of fatty acids is a useful tool in selection for various concentrations of fatty acids to suit consumer preferences (Hyten et al., 2004b; Pantalone et al., 2004).

Objectives

The objectives of this study were to:1) validate soybean agronomic and seed quality trait QTL in an independent Essex x Williams 82 RIL population; 2) report novel QTL for soybean agronomic and seed quality QTL and; 3) submit a manuscript to the Soybean Genetics Committee for review so that confirmed QTL may be named and published in Soybase.

References

References

Anderson, James W., Belinda M. Smith, and Carla S. Washnock. 1999. Cardiovascular and renal benefits of dry bean and soybean intake. Am. J. Clin. Nutr. 70:464 - 474.

Beavis, William D. 1998. QTL Analyses: Power Precision and Accuracy. p. 145-161. *In* Andrew H. Paterson (ed.) Molecular Dissection of Complex Traits. CRC Press LLC Boca Raton, FL.

Beavis, W. D., 1994. The power and deceit of QTL experiments: Lessons from comparative QTL studies. p. 250–266 *In* 49th Annual Corn and Sorghum Industry Research Conference. American Seed Trade Association, Washington, DC.

Bernard, R. L. and C.R. Cremeens. 1988. Registration of 'Williams 82' soybean Crop Sci 28:1027.

Bernardo, Rex. 2002. Breeding for Quantitative traits in Plants. p. 4, 292-294. Stemma Press, Woodbury, MN.

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

Fasoula, Vasilia A., Donna K. Harris and H. Roger Boerma. 2004. Validation and designation of quantitative trait loci for seed protein, seed oil, and seed weight from two soybean populations. Crop Sci. 44:1218-1225.

Friedman, M. and D. L. Brandon. 2001. Nutritional and health benefits of soy protein. J. Agric. Food Chem. 49:1069-86.

Hymowitz, T. and R. J. Singh. 1987. Taxonomy and Speciation. p. 25-48. *In* J.R. Wilcox (ed.) Soybeans: Improvement, Production, and Uses, 2nd ed. Agron. Monogr. 16. ASA, CSSA, SSSA, Madison, WI.

Hyten, D. L., V. R. Pantalone, C. E. Sams, A. M. Saxton, D. Landau-Ellis, T. R. Stefaniak, M. E. Schmidt. 2004a. Seed quality QTL in a prominent soybean population. Theor. Appl. Genet. 109:552-561.

Hyten, D. L., Vincent R. Pantalone, Arnold M. Saxton, Michael E. Schmidt and Carl E. Sams. 2004b. Molecular mapping and identification of soybean fatty acid modifier quantitative trait loci. J. Am. Oil Chem. Soc. 81:1115-1118.

Liu, KeShun. 1997. Soybeans: Chemistry, Technology and Utilization. p. 20-25. International Thompson Publishing, New York.

Lusas, Edmund W. 2004. Soybeans, processing and utilization. *In* H.R Boerma and J. E. Specht (eds.) Soybeans: Improvement, Production, and Uses, 3rd ed. Agron. Monogr. 16. ASA, CSSA, SSSA, Madison, WI.

Monsanto Company**.** Monsanto Launches Vistive™ Soybeans. [Online] Available at http://www.vistive.com/about.aspx verified 13 April 2008)

Pantalone, Vincent R., David R. Walker, Ralph Dewey and Istvan Rajcan. 2004. DNA Marker-Assisted Selection for Improvement of Soybean Oil Concentration and Quality. p. 284-285, 301-303. *In* H. Thomas Stalker and E. Charles Brummer (eds.) Legume and crop Genomics. AOCS Press, Champaign, Illinois.

Panthee, D.R., V.R. Pantalone, A.M. Saxton, D.R. West and C.E. Sams. 2006. Genomic regions associated with amino acid composition in soybean. Molecular Breeding 17: 79 – 89.

Panthee, D. R. V. R. Pantalone, D. R. West, A. M. Saxton, and C. E. Sams. 2005. Quantitative Trait Loci for Seed Protein and Oil Concentration, and Seed Size in Soybean. Crop Sci. 45: 2015-2022.

Prince, James and Ebenezer A. Ogundiwin. 2004. Molecular Mapping and Marker-Assisted Selection of Quantitative Trait Loci in Plants.p.129-151. *In* Paul Christou and Harry Klee (eds.) Handbook of Plant Biotechnology Vol. 1. John Wiley & Sons Ltd., West Sussex, England.

SAS Institute. 2002. The SAS system for Windows. Release 9.0. SAS Inst., Cary, NC.

Smith T.J.and H. M. Camper .1973. Registration of 'Essex' soybean. Crop Sci 13:495.

Sebolt, A.M., R.C. Shoemaker, and B.W. Diers 2000. Analysis of a Quantitative Trait Locus Allele from Wild Soybean That Increases Seed Protein Concentration in Soybean**.** Crop Sci. 40: 1438-1444.

Smith, Keith J. and Wipada Huyser 1987. World Distribution and Significance of Soybean p. 1-22. *In* J.R. Wilcox (ed.) Soybeans: Improvement, Production, and Uses, 2nd ed. Agron. Monogr. 16. ASA, CSSA, SSSA, Madison, WI.

Soyfoods.[online] Available at www.soyfoods.com (verified 14 April 2008).

Soystats. [online] Available at www.soystats.com (verified 14 April 2008).

Stefaniak, T. R., D. L. Hyten, V. R. Pantalone, A. Klarer and T.W. Pfeiffer. 2006. Soybean cultivars resulted from more recombination events than selected lines in the same population. Crop Sci. 46:43-51.

United State Department of Agriculture. Economic research Service. Soybean industry statistics. [Online] Available at (http://www.ers.usda.gov/News/soybeancoverage.htm verified 13 April 2008).

Wilcox J.R. 1998. Increasing seed protein in soybean with eight cycles of recurrent selection. Crop Sci. 38:1536-1540.

Wilcox, James R. 2004. World Distribution and Trade of Soybean. *In* H.R Boerma and J. E. Specht (eds.) Soybeans: Improvement, Production, and Uses, 3rd ed. Agron. Monogr. 16. ASA, CSSA, SSSA, Madison, WI.

Wilcox, James R., and Richard M. Shibles. 2001. Interrelationships among Seed Quality Attributes in Soybean. Crop Sci. 41:11-14.

Part 2

Detection and Validation of Agronomic Trait QTL in Soybean

Abstract

Soybean [Glycine max (L.) Merrill] seed yield is an important quantitative trait governed by many genes which may also correlate with other important traits, such as plant height, lodging and days to maturity. Despite the complex nature of quantitative trait expression, plant breeders have managed to make steady improvements using conventional and molecular breeding techniques. Genomic regions associated with quantitative traits, termed quantitative trait loci (QTL) have been published in Soybase. QTL validation studies can increase the usefulness of QTL in marker assisted selection (MAS) programs by eliminating erroneously reported QTL. The purpose of this study was to validate previously reported agronomic trait QTL in an independent 'Essex' x 'Williams 82' recombinant inbred line (RIL) population. Seventy SSR markers placed on linkage groups (LGs) B2, C2, D1a, D1b, D2, F, G, I, H, K, and M were analyzed by single factor ANOVA for possible association with least square means for seed yield, plant height, days to maturity and lodging in the confirmation population. Markers were selected based on being in intervals or flaking intervals previously reported to be QTL.

There were significant ($p < 0.05$) genotypic differences among the RILs for seed yield, plant height, maturity and lodging. RILs averaged 2596 kg ha^{-1} seed yield, 86cm plant height, 1.7 lodging score and 139 days to maturity. Agronomic trait QTL were confirmed as follows: seed yield [LG L (Satt166, Satt527, Satt561, and Satt229), maturity [LG C2 (Satt557, Satt460and Satt079), LG L (Satt156, Satt076, Satt166, Satt527, Satt561 and Satt229), plant height [LG C2 (Satt460, Satt371), LG L (Satt481, Satt156, Satt076,

Satt166, Satt527 and Satt561), LG M (Satt229)] and lodging [LG L (Satt076, Satt166, Satt527, Satt561, and Satt229)].

Thus, we were able to validate previously reported agronomic trait QTL to guide molecular breeders, and provide evidence for novel QTL that may affect further genetic gains in soybean.

Introduction

Soybean [Glycine max (L.) Merrill] agronomic traits such as seed yield, plant height, lodging and relative maturity, are quantitative traits whose inheritance is controlled by many genes with large or small effects. Direct selection of polygenic traits is difficult and is compounded by the variable expression of phenotypes across environments caused by the genotype x environment interaction. However, improvements have been made over the years using both conventional and molecular breeding techniques.

Specht et al. (1999) used annual soybean yield estimates from 1924 to 1998 to determine the rate of yield improvement in soybean. They reported that soybean yield improvement was 40% faster between 1972 and 1998 (about 31.2 ± 4.8 kg ha^{-1} yr^{-1}) than it was in the earlier part of the century. In another study, Wilcox (2001) assessed the genetic gain on yield over a 60 year period using data from the Northern Uniform Tests. There was a linear increase in yield for soybeans in maturity groups (MG) I-IV averaging 30 kg ha^{-1} yr^{-1}. The study also showed improved lodging resistance and small yet significant increases in plant height. Ustun et al. (2001) conducted a study to determine yield increases due to plant breeding efforts in the United States southern growing region. They reported annual gain in seed yield at 14 kg ha^{-1}, improved yield stability in fifth generation cultivars over that of ancestral lines and a general shift towards earlier maturing varieties. There was a decrease in plant height reported initially but, as lodging resistance improved, there was a shift towards taller varieties in the 1970s.

Molecular breeding techniques can improve the precision and reduce the time required to select for traits of interest. Many genomic regions associated with quantitative traits, termed quantitative trait loci (QTL), have been published in Soybase (http://www.soybase.org; verified 02 April 2008) and are available for breeders' use.

QTL detection studies are an integral part of marker-assisted selection breeding programs. Errors do occur in QTL studies resulting in erroneous association of QTL with a trait or a QTL not being detected at all. These errors may be reduced by conducting QTL confirmation studies. Fasoula et al. (2004) point out that due to lack of QTL confirmation studies. MAS breeding programs often have to conduct preliminary trials to validate existence of QTL before using reported and unconfirmed QTL in breeding experiments. Conducting confirmation experiments using independent populations and diverse environments to verify the existence of reported QTL can eliminate this initial step (Fasoula et al., 2004).

The Soybean Genetics Committee has established criteria which must be satisfied before a QTL can be confirmed. The QTL must be tested in new meiotic events and different environments than those it was detected in. The confirmation population cross must contain at least one of the parents from the original mapping study or be derived from an individual from the original mapping study that is segregating at the locus of interest and an alpha level of 0.01 or less must be used to test the hypotheses (http://soybase.org/resources/QTL.php ; verified 02 April 2008).

Fasoula et al. (2004) confirmed two protein QTL cqProt-001 (E/A454-1) and cqProt-002 (UNK/A132-4) three oil QTL cqOil-001 (C1/A063-1), cqOil-002(H/A566-2), cqOil-003 (L/Satt398) and two seed weight QTL cqSd wt-001(G/Satt303), cqSd wt-002

(E/Satt263). Where the first letter within the parenthesis indicates the LG and the second the RFLP marker. The LG denoted UNK was unlinked to known linkage groups.
In another study, Nichols et al. (2006) confirmed QTL for seed yield (cqSd yld -001), maturity (cqPod mat-001) and seed size (cqSd wt-003) on LG I. Confirmation of the numerous agronomic trait QTL reported in Soybase would provide breeders with a marker assisted selection tool for cultivar improvement.

Hyten (2002) mapped several agronomic trait QTL in an 'Essex' x 'Williams' recombinant inbred line (RIL) population. Essex and Williams belong to the Southern and Northern germplasm pools respectively, and are the parents of the prominent line 'A3127'. A3127 has been widely used in the pedigrees of many elite lines both in the Northern, and the Southern regions of the United States (Sneller, 1994). Therefore QTL confirmed in our population which is derived from an Essex x Williams 82 cross may be useful to many plant breeders because Williams 82 is a near isogenic line (NIL) of Williams..

The purpose of this study was to validate those QTL in an independent Essex x 'Williams 82' RIL population, report novel QTL for seed yield, plant height, maturity and lodging and to submit a manuscript to the Soybean Genetics Committee for review so that confirmed QTL may be named and published in Soybase.

Materials and Methods

Essex x Williams 82 Population Development

The initial crosses for the Essex x Williams 82 population were made at the University of Tennessee Plant Science Farm (KPSF) of the East Tennessee Research and

Education Center (ETREC) in Knoxville, TN in the summer of 2002. The population was advanced from the F_2 to the F_6 generation through single seed descent (Brim, 1966)

Field Experiments

In 2006, 146 recombinant inbred lines (RILs) and the two parents, Essex and Williams 82, were planted in four reps of single plant hill plots (sown 3 beans/hill then thinned to one plant per plot about 14 days after emergence) in a randomized complete block Honeycomb design (Fasoulas and Fasoula, 1995). In addition, 129 genotypes were planted in 3.1m rows sown at 8 beans/0.31m for seed increase for the 2007 experiments. Lines which did not yield enough seed were further increased in winter nursery to ensure that there was sufficient seed for the 2007 experiments.

Based on maturity data collected in 2006, the population was divided by days to maturity into three population subsets: early (22 genotypes), mid (96 genotypes) and late (23 genotypes) separated from each other by ten days in maturity. Five genotypes were dropped from the study because of inadequate seed amounts. In 2007, each population subset which included the two parents was planted in two 6.1m row plots in a randomized complete block design replicated three times at the East Tennessee Research and Education Center (ETREC), the Highland Rim Research and Education Center (HRREC) and the West Tennessee Research and Education Center (WTREC). Checks in the population subsets were assigned by maturity group as follows: Early (‘Macon’ and ‘LD00-3309’), Mid (‘5002T’) and Late (5002T and ‘5601T’).

Phenotypic Traits

Flower color was recorded when more than 95% of the plants in a plot were in full bloom (R_2). Maturity be recorded when 95% of the pods in a plot showed their mature color, at R_8 growth stage. Pubescence color was also recorded at maturity. Maturity was calculated as the number of days from date of planting to R_8. Plant height was measured as the average row height from the base of the stem at the soil surface to the top of the plant at maturity within a plot (Fehr et al., 1971). Lodging was scored at maturity with a score range from 1 (all plants erect) to 5 (all plants prostrate). Plots were end-trimmed to 4.9 m prior to harvest. Yield in lbs per plot and percent moisture were recorded at harvest. For data analysis yield per plot was adjusted to 13% moisture and converted to kg ha^{-1}.

DNA Extraction and Polymerase Chain Reaction

In 2006, four plants each of 146 RILs and the two parents were grown in the growth chamber for leaf collection for DNA isolation. Leaves were harvested from all four plants at V4 stage. In addition, five leaves per genotype were randomly sampled from the first replication of plots in each test in ETREC in 2007 to ensure that we had enough leaves from which to extract DNA. All leaves were flash frozen in liquid nitrogen and then stored at -80 ºC until DNA was extracted.

DNA was extracted from the leaves of RILs and parents using the Qiagen Plant DNeasy Extraction Kit (Qiagen, Valencia, CA). Polymerase chain reaction (PCR) consisted of 6.9 µL of ddH_2O, 4 µL Hotmaster TAQ (5Prime Gaithersburg, MD), 0.55 µL of 5 µM WellRED labeled forward primer (Proligo, Boulder, CO) and 0.55 µL of

5 µM reverse primer (Sigma-Genosys, Woodlands, TX), and 3 µL of 15 ng μL^{-1} template DNA. PCR reactions were performed in a 96-well MBS Hybaid thermocycler (Hybaid, Franklin, MA). PCR conditions consisted of initial denaturation at 95ºC for 2 min followed by 35 cycles of: a) denaturation at 92ºC for 1 min, b) 47ºC for annealing for 1 min, c) 68ºC for 5 min for extension and, d) a final cycle at 72ºC for extension for 5 min. Parents were screened with a total of 203 $(ATT)_n$ type simple sequence repeat (SSR) genetic markers developed by Cregan et al. (1999) with the goal to use markers to confirm agronomic trait QTL. One hundred and seventeen markers were found to be polymorphic. Seventy of 117 polymorphic SSR markers which amplified in the progeny were used to genotype the RILs for QTL analysis. Sequence information for the SSRs is publicly available from Soybase (www.soybase.org; verified April 3, 2008.)

DNA Gel Electrophoresis

A Beckman-Coulter CEQ 8800 Genome Analysis System (Beckman-Coulter, Fullerton, CA) was used to separate PCR products by gel capillary electrophoresis. Beckman Coulter fragment analysis software was used to estimate PCR amplicon base pair lengths for detectable polymorphisms. Markers were considered to be polymorphic when the parental amplicon base pair lengths differed by ≥3 nucleotides. Polymorphic primers detected in the parents were used to genotype RILs. RILs were scored (1 = Essex allele, 2 = Heterozygote, and 3 = Williams 82 allele) for all markers. Heterozygote scores were dropped from the data set so as to only report potential additive genetic effects.

Data Analysis

The early, mid, and late population subsets were combined across the three environments (ETREC, HRREC and WTREC) to create three populations named set-1, set-2 and set-3 respectively. It was hypothesized that by separating the populations by maturity group we might be able to see QTL differences within a narrower range of maturity. Sets 1, 2 and 3 were also combined to form the total population which could enable us to further detect QTL with greater power.

Phenotypic data were analyzed using the MIXED procedure of SAS (SAS Institute Inc., 2002) software to determine whether there were significant genotypic differences among lines, within sets and within the total population. Environments and replications were considered as two blocking factors in the model.

QTL were analyzed via single factor ANOVA GLM (SAS Institute Inc., 2002) procedure to determine if there were any associations between SSR markers and phenotypic least square means for the traits of interest. MAPMAKER /EXP 3.0 (Lander et al.1987; Lincoln et al.1992) was used to estimate linkage distance between SSR markers. A minimum likelihood of odds (LOD) $\geq$ 3.0 and a maximum distance $\leq$ 50 cM were used to test linkage among markers. Marker order was similar to the soybean composite integrated map (Song et al,. 2004), a finding also supported in the original Essex x Williams population by Hyten et al. (2004a). The soybean composite integrated map was used to determine marker distances because it is based on five mapping populations and marker distances are more accurate.

Heritability for each trait was calculated on an entry mean basis for three environments and three replications according to Nyquist (1991) whose formula was presented by Panthee et al. (2005).

REML estimation in PROC MIXED was used to determine the variance components for heritability calculations. PROC CORR (SAS 2002) was used to determine phenotypic correlations between traits.

Results and Discussion

Seventy SSR markers placed on linkage groups (LGs) B2, C2, D1a, D1b, D2, F, G, I, H, K, and M were analyzed by single factor ANOVA in the Essex x Williams 82 population for their possible association with least square means for seed yield, plant height, days to maturity and lodging. Markers were selected based on being in intervals, or flaking intervals, previously reported by Hyten (2002) and Hyten et al. (2004a and 2004b) to be QTL.

There were significant ($p < 0.05$) genotypic differences among the RILs within all sets and within the total population for plant height, maturity and lodging. Genotypic differences among RILs seed yield were significant ($p<0.05$) within set-1, set-2 and the total population. QTL analyses for seed yield were not performed for set-3 because genotypic differences among the RILs were not significant ($p > 0.05$).

Seed yield for the total RIL population ranged from 1530 to 3492 kg ha^{-1}, with a mean of 2596 kg ha^{-1} (Table 2.1).

Table 2.1 Descriptive statistics for agronomic traits in an F_6 derived soybean RIL population from an Essex x Williams 82 cross grown in three environments in Tennessee

Trait	Min	Max	Mean	Std. Dev	Essex	Williams 82	h^2
Yield (kg ha^{-1})	1530	3412	2596	339.6	2999	1715	0.45
Maturity (days)	121	147	139	5.1	142	127	0.95
Height (cm)	56	125	86	15.0	80	84	0.94
Lodging[a]	1.0	3.3	1.7	0.5	1.6	2	0.17

[a]Lodging scored from 1 (all plants erect) to 5(all plants prostrate)

Table 2.2 Phenotypic correlation coefficients among agronomic traits in an F_6 derived soybean RIL population from an Essex x Williams 82 cross grown in three environments in Tennessee

Trait	Height	Lodging	Maturity
Yield	-0.08 ^{ns}	-0.33****	0.40****
Height		0.57****	0.36****
Lodging			0.10 ^{ns}

*, **, ***, ****Significant at P=0.05, P=0.01, P=0.001and P=0.0001; ns, non-significant

Even though there was a large seed yield difference between the parents, the wide range observed in the RILs was due to transgressive segregation (based on ± 2 std. deviations from the high parent or low parent). There was a significant positive phenotypic correlation between seed yield and maturity ($r= 0.40$, $p < 0.0001$) and a significant negative phenotypic correlation with lodging ($r= -0.33$, $p < 0.0001$) (Table 2.2). Seed yield was moderately heritable and was not significantly correlated with plant height.

Hyten (2002) detected yield QTL on LGs A1 and L in an Essex x Williams RIL population. In the confirmation study, no markers were placed on A1, but since it was a novel QTL in the original mapping study, work may be done to validate it in the future. A QTL explaining 13% of yield variation was detected on LG L within the interval Satt166-Satt229 in one environment in the original mapping study. In the confirmation study, Y10, Y11, Y12 and Y13 in that interval were confirmed ($p < 0.01$) in set-1 and Y10, Y11, Y12 in the total population (Table 2.3). In the total population, Y13 was detected but not confirmed because its p-value (0.012) exceeds the confirmation threshold of $p < 0.01$. No yield QTL in this region of LG L were detected in population set-2. Other yield QTL on LG L detected in this study, 5 cM upstream from the map position reported by Hyten (2002), were: Y7 (set-1 and the total population), Y8 (set-2 and the total population) and Y9 (set-1 and the total population). Mapping with MAPMAKER (Lander et al., 1987) showed these two intervals to be linked in this study (data not shown).

Other yield QTL not reported by Hyten (2002) were detected on LGs C2, D1a, D1b, D2 and M. The yield QTL, Y1 (set-1 and set-2), Y2 (set-2) and Y3 (set-3) on LG C2 were considered to be major QTL (R^2 >10%), and they confirmed QTL in the interval

Table 2.2 Phenotypic correlation coefficients among agronomic traits in an F_6 derived soybean RIL population from an Essex x Williams 82 cross grown in three environments in Tennessee

Trait	Height	Lodging	Maturity
Yield	-0.08 ^{ns}	-0.33****	0.40****
Height		0.57****	0.36****
Lodging			0.10 ^{ns}

*, **, ***, ****Significant at P=0.05, P=0.01, P=0.001and P=0.0001; ns, non-significant

Table 2.3 QTL, map position and additive genetic effects for seed yield in F_6 derived early, medium and late maturing soybean RIL subpopulations and total population from an Essex x Williams 82 cross grown in three environments in Tennessee

				Set-1			Set-2			Set-3			Total Population		
QTL[a]	Marker[b]	LG[c]	Pos[d]	Effect[e]	P	R^2	Effect[e]	P	R^2	Effect[e]	P	R^2	Effect[e]	P	R^2
Y1	Satt557	C2	112.19	759	0.0001	0.41	-	ns	-	-	ns	-	528	0.0001	0.18
Y2	Satt079	C2	117.87	431	0.0079	0.11	-	ns	-	-	ns	-	-	ns	-
Y3	Satt460	C2	117.77	599	0.0001	0.25	-	ns	-	-	ns	-	279	0.0003	0.10
Y4	Satt129	D1a	11.22	-492	0.0052	0.17	-	ns	-	-	ns	-	-	ns	-
Y5	Satt266	D1b	59.61	369	0.0397	0.08	-	ns	-	-	ns	-	166	0.0126	0.06
Y6	Satt372	D2	39.35	-	ns	ns	165	0.0395	0.06	-	ns	-	-	ns	-
Y7	Satt481	L	54.57	462	0.0049	0.14	-	ns	-	-	ns	-	202	0.0008	0.09
Y8	Satt156	L	56.14	ns	ns	ns	218	0.0093	0.03	-	ns	-	258	0.0001	0.14
Y9	Satt076	L	61.35	676	0.0001	0.32	-	ns	-	-	ns	-	213	0.0002	0.11
Y10	Satt166	L	66.50	719	0.0001	0.34	-	ns	-	-	ns	-	237	0.0001	0.12
Y11	Satt527	L	70.36	753	0.0001	0.37	-	ns	-	-	ns	-	227	0.0002	0.11
Y12	Satt561	L	71.44	716	0.0001	0.35	-	ns	-	-	ns	-	247	0.0002	0.13
Y13	Satt229	L	93.89	564	0.0019	0.18	-	ns	-	-	ns	-	155	0.0117	0.06
Y14	Satt590	M	7.84	-	ns	-	214	0.0060	0.03	-	ns	-	196	0.0023	0.08
Y15	Satt540	M	35.85	-	ns	-	264	0.0072	0.03	-	ns	-	-	ns	-
Y16	Satt245	M	53.55	-	ns	-	164	0.0066	0.03	-	ns	-	237	0.0002	0.11
Y17	Satt175	M	66.99	-	ns	-	ns	ns	ns	-	ns	-	206	0.0033	0.08

[a] Y1 to Y17 denote seventeen seed yield QTL detected

[b] SSR marker associated with the QTL by SF ANOVA

[c] Linkage group based on the integrated soybean genetic linkage map

[d] Marker position on linkage group based on the integrated soybean genetic linkage map

[e] Average change (kg ha^{-1}) with respect to Essex allele

Satt307-Satt079 detected by Kassem et al. (2006) in an Essex x 'Forrest' RIL population. Even though Y1, Y2 and Y3 are confirmed yield QTL with large effects, they were associated with seed yield in fields showing sudden death syndrome (SDS) symptoms (Kassem et al., 2006). Moreover, mild SDS symptoms were observed in ETREC and WTREC, so selection for these QTL may introduce SDS susceptibility. On LG D1a, Y4 (set-1) is 10 cM upstream from the closest seed yield QTL reported in Soybase and may be a new QTL. Y5 (set-1 and the total population), Y6 (set-2), on LGs D1b and D2 respectively, have previously been reported (Orf et al., 1999). Y14 on LG M is a new QTL detected in this study explaining 2% and 8% yield variation (set-2 and set-3, respectively) and is 10 cM upstream from another yield QTL reported by Orf et al. (1999). Y15 (set-2) and Y16 (set-2 and set-3) also on LG M have been previously reported (Specht et al., 2001). Y17 has an R^2 of 8% and is 16cM down stream from the closest yield QTL listed in Soybase and thus may be a new QTL.

There was a difference in days to maturity between parents and RILs (Table 2.1). Maturity QTL reported by Hyten (2002) and other maturity QTL were detected (Table 2.4). M1 and M2 detected in LG B2 are new QTL not reported in Soybase, each explaining 7% of maturity variation. M1 and M2 lie very close to each other (< 1 cM) so this is probably a single QTL. M3 was detected but not confirmed ($P > 0.01$) in the total population only. M4 was confirmed in sets 1, 3, and the total population, while M5 and M6 were both confirmed in sets 1 and 2. M3 –M6 lie within the same interval as the E1 gene (Cober and Voldeng, 2001) which would explain the high R^2 values observed in this population. Moreover Y1 resides at the same locus as M4, and Y3 resides at the same locus as M5. M7 and M8 on LG D1a are about 26 cM upstream from the closest maturity

Table 2.4 QTL map position and additive genetic effects for days to maturity in F_6 derived early, mid and late maturing soybean RIL subpopulations and total population from an Essex x Williams 82 cross grown in three environments in Tennessee

				Set-1			Set-2			Set-3			Total Pop.		
QTL[a]	**Marker[b]**	**LG[c]**	**Pos[d]**	**Effect[e]**	**P**	**R^2**	**Effect[e]**	**P**	**R^2**	**Effect[e]**	**P**	**R^2**	**Effect[e]**	**P**	**R^2**
M1	Satt070	B2	72.8	-	ns	-	-	ns	-	3	0.0407	0.07	-	ns	-
M2	Satt474	B2	75.34	-	ns	-	-	ns	-	4	0.0339	0.07	-	ns	-
M3	Satt144	C2	112.08	-	ns	-	-	ns	-	-	ns	-	-2	0.0492	0.03
M4	Satt557	C2	112.19	8	0.0001	0.35	-	ns	-	11	0.0033	0.21	10	0.0001	0.43
M5	Satt460	C2	117.77	8	0.0001	0.30	-	ns	-	-	ns	-	5	0.0001	0.13
M6	Satt079	C2	117.87	6	0.0007	0.17	-	ns	-	-	ns	-	3	0.0041	0.07
M7	Satt129	D1a	11.22	-6	0.0042	0.17	-	ns	-	-	ns	-	-	ns	-
M8	Satt147	D1a	11.99	-5	0.005	0.15	2	0.0464	0.02	-	ns	-	-	ns	-
M9	Satt240	K	52.88	-	ns	-	2	0.0261	0.02	-	ns	-	-	ns	-
M10	Satt481	L	54.57	8	0.0001	0.26	-	ns	-	-	ns	-	-	ns	-
M11	Satt156	L	56.14	7	0.0037	0.19	-	ns	-	-	ns	-	-	ns	-
M12	Satt076	L	61.35	9	0.0001	0.44	-	ns	-	-	ns	-	-	ns	-
M13	Satt166	L	66.55	10	0.0001	0.45	-	ns	-	-	ns	-	-	ns	-
M14	Satt527	L	70.36	9	0.0001	0.40	-	ns	-	-	ns	-	-	ns	-
M15	Satt561	L	71.44	9	0.0001	0.37	-	ns	-	-	ns	-	-	ns	-
M16	Satt229	L	93.89	5	0.0085	0.13	-	ns	-	-	ns	-	-	ns	-
M17	Satt540	M	35.85	-	ns	-	2	0.0418	0.02	-	ns	-	-	ns	-

[a] M1 to M17 denote seventeen maturity QTL detected
[b] SSR marker associated with the QTL by SF ANOVA
[c] Linkage group based on the integrated soybean genetic linkage map
[d] Marker position on linkage group based on the integrated soybean genetic linkage map
[e] Average change (days to maturity) with respect to Essex allele

QTL published in Soybase (Keim et al., 1990). M9 detected on LG K in this population was also detected in other studies (Lee et al., 1996).Maturity QTL on LG L were confirmed (M11-M16) in set-1. M10 was detected and lies 2 cM upstream from M11, so this region may constitute a single QTL. Satt229 (M16) was mapped by Hyten (2002) in the same interval as M10-M15; however in this study it was unlinked to those markers, perhaps because there were no polymorphic markers available in the confirmation population. Maturity QTL have been mapped close to M16 (Orf et al., 1999). M17 was detected on LG M ($p < 0.05$), but was not confirmed. The limited number of markers found to be polymorphic in this interval may be the reason why no QTL were confirmed on this LG. Maturity QTL associated with maturity on LGs D1b, F and J reported by Hyten (2002) were not detected in this study. The large heritability estimate associated for maturity may be explained by the large number of major maturity QTL (R^2 ranging from 13% to 44%) detected on LGs C2 and L.

Plant height among RIL in the total population ranged from 56 to 125 cm, with a mean of 86 cm. There was a positive correlation between height and lodging ($r = 0.57$, $p < 0.0001$) and height and maturity ($r = 0.36$, $p < .0001$) (Table 2.2).

Hyten (2002) reported plant height QTL on LGs C2, F and L. H4 and H7 on LG C2 were confirmed in set-1 and set-2 respectively. H3 (set-1), H5 (set-1) and H6 were detected but not confirmed ($p > 0.01$) (Table 2.5). Plant height QTL on LG F, H15-H17, were detected but not confirmed ($0.01 < p < 0.05$). H15 is about 45 cM upstream from H16, so it is likely a different QTL on LG F. H27 on LG L was confirmed in sets 1 and 2 and detected but not confirmed in the total population $0.01 < p < 0.05$. H28-H32 were confirmed on LG

Table 2.5 QTL, map position and additive genetic effects for plant height in F_6 derived early, medium and late maturing soybean RIL subpopulations and total population from an Essex x Williams 82 cross grown in three environments in Tennessee

				Set-1			Set-2			Set-3			Total Population		
QTL[a]	**Marker[b]**	**LG[c]**	**Pos[d]**	**Effect[e]**	**P**	**R^2**	**Effect[e]**	**P**	**R^2**	**Effect[e]**	**P**	**R^2**	**Effect[e]**	**P**	**R^2**
H1	Satt070	B2	72.8	-	ns	-	-6	0.0209	0.02	-	ns	-	-	ns	-
H2	Satt474	B2	75.34	-	ns	-	-12	0.0455	0.02	-	ns	-	-	ns	-
H3	Satt557	C2	112.19	-9	0.0287	0.12	-	ns	-	-	ns	-	-	ns	-
H4	Satt460	C2	117.77	-12	0.0009	0.18	-	ns	-	-	ns	-	-	ns	-
H5	Satt079	C2	117.87	-7	0.0438	0.07	-	ns	-	-	ns	-	-	ns	-
H6	Satt202	C2	126.24	-	ns	-	-6	0.0248	0.02	-	ns	-	-	ns	-
H7	Satt371	C2	145.48	-	ns	-	-8	0.0042	0.04	-	ns	-	-7	0.0284	0.05
H8	Satt147	D1a	11.99	10	0.0040	0.15	7	0.0096	0.03	-	ns	-	-	ns	-
H9	Satt203	D1a	61.89	ns	ns	ns	7	0.0091	0.03	-	ns	-	-	ns	-
H10	Satt179	D1a	64.69	7	0.0397	0.07	8	0.0016	0.04	-	ns	-	6	0.0302	0.04
H11	Satt172	D1b	100.89	-	ns	-	-9	0.0026	0.05	-	ns	-	-6	0.0349	0.04
H12	Satt274	D1b	116.35	-	ns	-	-5	0.0400	0.02	-	ns	-	-	ns	-
H13	Satt459	D1b	118.62	-	ns	-	-5	0.0500	0.02	-	ns	-	-	ns	-
H14	Satt372	D2	39.35	-	ns	-	-7	0.0041	0.03	-	ns	-	-7	0.0250	0.05
H15	Satt348	F	15.29	-9	0.0150	0.11	-6	0.0300	0.02	-	ns	-	-6	0.0426	0.03
H16	Satt114	F	63.69	-7	0.0466	0.08	-	ns	-	-	ns	-	-	ns	-
H17	Satt335	F	77.70	-7	0.0489	0.07	-	ns	-	-	ns	-	-	ns	-
H18	Satt144	F	102.08	-13	0.0004	0.20	-	ns	-	-	ns	-	-5	0.0317	0.04
H19	Satt522	F	119.19	-7	0.0393	0.08	-	ns	-	-	ns	-	-	ns	-

Table 2.5 Continued

					Set-1			Set-2			Set-3		Total population		
QTL[a]	**Marker[b]**	**LG[c]**	**Pos[d]**	**Effect[e]**	**P**	**R^2**	**Effect[e]**	**P**	**R^2**	**Effect[e]**	**P**	**R^2**	**Effect[e]**	**P**	**R^2**
H20	Satt427	G	51.69	11	0.0057	0.13	-	ns	-	-	ns	-	-	ns	-
H21	Satt192	H	44.04	12	0.0033	0.15	-	ns	-	-	ns	-	-	ns	-
H22	Satt539	K	1.798	-	ns	-	6	0.0100	0.03	-	ns	-	-	ns	-
H23	Satt240	K	52.879	-	ns	-	8	0.0044	0.03	-	ns	-	-	ns	-
H24	Satt725	K	56.85	8	0.0169	0.09	-	ns	-	-	ns	-	-	ns	-
H25	Satt475	K	78.68	7	0.0444	0.07	-	ns	-	-	ns	-	-	ns	-
H26	Satt260	K	80.119	10	0.0135	0.12	-	ns	-	-	ns	-	-	ns	-
H27	Satt481	L	54.574	-12	0.0010	0.18	-7	0.0066	0.03	-	ns	-	-7	0.0255	0.04
H28	Satt156	L	56.14	-12	0.0022	0.21	-12	0.0001	0.08	-	ns	-	-10	0.0016	0.09
H29	Satt076	L	61.35	-13	0.0002	0.22	-19	0.0001	0.23	-	ns	-	-15	0.0001	0.25
H30	Satt166	L	66.55	-14	0.0001	0.24	-23	0.0001	0.29	-	ns	-	-19	0.0001	0.37
H31	Satt527	L	70.36	-12	0.0004	0.22	-22	0.0001	0.28	-	ns	-	-18	0.0001	0.37
H32	Satt561	L	71.44	-9	0.0042	0.13	-23	0.0001	0.32	-	ns	-	-18	0.0001	0.37
H33	Satt229	L	93.885	-	ns	-	-19	0.0001	0.21	-	ns	-	-16	0.0001	0.27

[a] H1 to H33 denote thirty three plant height QTL detected
[b] SSR marker associated with the QTL by SF ANOVA
[c] Linkage group based on the integrated soybean genetic linkage map
[d] Marker position on linkage group based on the integrated soybean genetic linkage map
[e] Average change (cm) with respect to Essex allele

L in all population sets and in the total population. H33 was confirmed in sets 2 and in the total population. Large effects of the height QTL detected on LG L are probably due to the proximity of those loci to the stem terminency Dt1 locus. Other studies have also found this region to be significantly associated with plant height (Specht at al., 2001; Lee et al., 1995).

Plant height QTL detected on LGs B2, D1a, D1b, D2, G, H and K were not detected in the original mapping study. H1 and H2 detected in set-2 on LG B2 have small effects (R^2 2%) and have not been previously reported on Soybase. H1 and H2 are < 1cM apart so they probably represent a single QTL. Plant height QTL on LG D1a were detected as follows: H8 in sets 1 and 2, H9 in set-2 and H10 in sets 1, 2 and the total population.H8 is 50 cM upstream from the H9-H10 interval which is about 20 cM upstream from the closest plant height QTL reported in Soybase. This suggests two new QTL on LG D1a. Plant height QTL on D1b were detected as follows: H11 in sets 2 and the total population, and H12 and H13 in set-2. These QTL are about 40 cM apart from the closest reported QTL in Soybase and may represent a new height QTL. There are no QTL for plant height reported in LG D2, so H14 may be another new QTL. QTL detected on LG G, H and K, with the exception of H23 and H24 (Yuan et al., 2002) are also likely to be new in this population as they are approximately 15 cM away from the closest plant height QTL reported in Soybase. The detection of many QTL near the Dt1 locus on LG L may explain the very high heritability estimate (0.95) calculated for this population.

Essex and Williams 82 had similar lodging scores (Table 2.1) while RIL lodging scores ranged from 1 to 3.3. Hyten (2002) detected lodging QTL on LGs C2, D1b, F and L. Lodging QTL were detected in the confirmation population on LGs C2, D1b, D2, F, I, K , L

and M (Table 2.6). Lodging QTL previously mapped on LGs C2, D1b and F were detected but not confirmed ($p > 0.01$). QTL previously mapped on LGs D1b by Hyten (2002) were new QTL in the original mapping study and they were detected but not confirmed in this study ($p > 0.01$). Our inability to confirm QTL on LGs C2, D1b, D2 and F maybe due to the scarcity of markers flanking those QTL in our population, or because they were not stable across environments in the mapping study and have a large genotype x environment interaction. L8, (set-1 and set-3) and L9 (total population) on LGs I and K were new QTL detected in the confirmation population which have not been reported in Soybase. L11–L12 and L13-15 on LG L were confirmed in set 2 and the total population. L12 and L13 were also detected in set-1. The interval on LG L spanning L11 to L14 is very close to the Dt1 locus that is also associated with height and lodging (Soybase). L16 detected on LG M was detected in the confirmation population but not in the mapping population and it explained 9% of phenotypic variation for lodging. There are no lodging QTL on LG M currently reported in Soybase, so L16 is a new QTL detected in this study.

Conclusion

Several QTL intervals conditioning agronomic traits (yield , height, maturity and lodging) reported by Hyten (2002) have been confirmed in this study by single factor ANOVA of least square mean associations with SSR markers (e.g. seed yield QTL Y10-Y13, maturity QTL M4-M6, height QTL H27-H32 and lodging QTL L13-L15). There were several QTL detected in this study that were previously not reported by Hyten (2002) perhaps because they may have been environment specific.

Table 2.6 QTL, map position and additive genetic effects for lodging in F_6 derived early, medium and late maturing soybean RIL subpopulations and total population from an Essex x Williams 82 cross grown in three environments in Tennessee

				Set-1			Set-2			Set-3			Total Population		
QTL[a]	**Marker[b]**	**LG[c]**	**Pos[d]**	**Effect[e]**	**P**	**R^2**	**Effect[e]**	**P**	**R^2**	**Effect[e]**	**P**	**R^2**	**Effect[e]**	**P**	**R^2**
L1	Satt460	C2	117.77	-	ns	-	-	ns	-	-	ns	-	-0.2	0.0454	0.03
L2	Satt266	D1b	59.61	-	ns	-	-0.2	0.0123	0.03	-	ns	-	-	ns	-
L3	Satt412	D1b	72.57	-	ns	-	-	ns	-	-	ns	-	0.3	0.0170	0.09
L4	Satt014	D2	29.56	0.5	0.0244	0.08	-	ns	-	-	ns	-	-	ns	-
L5	Satt372	D2	39.35	-	ns	-	-0.2	0.013	0.03	-	ns	-	-	ns	-
L6	Satt649	F	5.36	0.5	0.0222	0.13	-	ns	-	0.6	0.0411	0.07	0.3	0.0062	0.07
L7	Satt269	F	11.37	-	ns	-	-	ns	-	0.5	0.0388	0.08	0.3	0.0395	0.05
L8	Satt292	I	82.78	0.4	0.0485	0.06	-	ns	-	-0.6	0.0374	0.08	-	ns	-
L9	Satt260	K	80.12	-	ns	-	-	ns	-	-	ns	-	0.2	0.0320	0.04
L10	Satt143	L	30.15	-	ns	-	-0.2	0.0460	0.02	-	ns	-	-0.3	0.0135	0.05
L11	Satt076	L	61.35	-	ns	-	-0.5	0.0001	0.07	0.6	0.0406	0.07	-0.2	0.0021	0.08
L12	Satt166	L	66.55	-0.4	0.0493	0.07	-0.3	0.0109	0.03	-	ns	-	-0.4	0.0001	0.15
L13	Satt527	L	70.36	-0.4	0.0375	0.08	-0.4	0.0008	0.05	-	ns	-	-0.4	0.0001	0.16
L14	Satt561	L	71.44	-	ns	-	-0.4	0.0003	0.06	-	ns	-	-0.4	0.0001	0.15
L15	Satt229	L	93.89	-	ns	-	-0.5	0.0001	0.09	-	ns	-	-0.4	0.0001	0.19
L16	Satt245	M	53.54	-	ns	-	-	ns	-	-0.6	0.0261	0.09	-	ns	-

[a] L1 to L16 denote sixteen lodging QTL detected

[b] SSR marker associated with the QTL by SF ANOVA

[c] Linkage group on the integrated soybean genetic linkage map

[d] Marker position on linkage group based on the integrated soybean genetic linkage map

[e] Average change [1= (all plants erect] to 5 (all plants prostrate)] with respect to Essex

Several QTL associated with all agronomic traits discussed were detected close to chromosomal regions known to control maturity in soybean (E1 gene and Dt1 locus). It is also interesting to note that in comparison to QTL detected in set-1 and in the total population, very few QTL were detected in sets 2 and 3. This may indicate that earliness has an effect on all the traits and that in the overall population there were other genotypic and environmental factors not related to maturity that affected trait expression. It would be interesting to further investigate the effects of maturity for QTL confirmation studies as it seems to have significant influence on important quantitative traits. Indeed, the Soybean Genetics Executive Committee suggested that research to precisely identify and locate maturity genes would greatly impact soybean improvement (Pantalone, personal communication, 2008).

References

References

Brim, C. A., 1966. A modified pedigree method of selection in soybeans. Crop Sci. 6: 220.

Cober, E.R., and H.D. Voldeng. 2001. A New soybean maturity and photoperiod-sensitivity locus linked to *E1* and *T*. Crop Sci. 41:698-701.

Cregan, P. B., T. Jarvik, A. L. Bush, R. C. Shoemaker, K. G. Lark, A. L. Kahler, N. Kaya, T. T. VanToai, D. G. Lohnes, and J. Chung, 1999. An integrated genetic linkage map of the soybean genome. Crop Sci. 39:1464-1490.

Fasoula, Vasilia A., Donna K. Harris and H. Roger Boerma. 2004. Validation and designation of quantitative trait loci for seed protein, seed oil, and seed weight from two soybean populations. Crop Sci. 44: 1218-1225.

Fasoulas, A.C and V.A Fasoula 2005. Honey comb selection designs. *In* Jules Janick (ed.) Plant Breeding Reviews. Vol. 13. John Wiley and Sons Inc. New York, USA.

Fehr, W.R., C.E. Caviness, D.L. Burmon and J.S. Pennington. 1971. Stages of development decriptions for soybeans for soybeans, Glycine max (L.) Merrill. Crop Sci. 11: 929-931.

Hyten, D. L. QTL mapping and identification of GxE interactions of agronomic and seed quality Traits in soybean. 2002.

Hyten, D. L., V. R. Pantalone, C. E. Sams, A. M. Saxton, D. Landau-Ellis, T. R. Stefaniak, and M. E. Schmidt, 2004a. Seed quality QTL in a prominent soybean population. Theor. Appl. Genet. 109: 552—561.

Hyten, D. L., Vincent R. Pantalone, Arnold M. Saxton, Michael E. Schmidt and Carl E. Sams. 2004b. Molecular mapping and identification of soybean fatty acid modifier quantitative trait loci. J. Am. Oil Chem. Soc. 81 (12):1115-1118.

Keim, P., B. W. Diers, T. C. Olson, and R. C. Shoemaker, 1990. RFLP mapping in soybean. association between marker loci and variation in quantative traits. Genetics 126: 735—742.

Lee, S. H., M. A. Bailey, M. A. R. Mian, E. R. Shipe, D. A. Ashley, W. A. Parrott, R. S. Hussey, and H. R. Boerma, 1996. Identification of quantitative trait loci for plant height, lodging, and maturity in a soybean population segregating for growth habit. Theor. Appl. Genet. 92: 516-523.

Lincoln, S. E., M. J. Daly, and E. S. Lander, 1993. Constructing Genetic Linkage Maps with MAPMAKER, Version 3. A Tutorial and Reference Manual. Release 3rd ed. Whitehead Institute for Biomedical Research, Cambridge, MA.

Nyquist, W. E., 1991. Estimation of heritability and prediction of selection response in plant populations. Crit. Rev. Plant Sci. 10:235-322.

Orf, J. H., K. Chase, T. Jarvik, L. M. Mansur, P. B. Cregan, F. R. Adler, and K. G. Lark, 1999b. Genetics of soybean agronomic traits. I. Comparison of three related recombinant inbred populations. Crop Sci. 39:1642-1651.

Pantalone, Vincent R., David R. Walker, Ralph Dewey and Istvan Rajcan. 2004. DNA Marker-Assisted Selection for Improvement of Soybean Oil Concentration and Quality. p. 284-285, 301-303. *In* H. Thomas Stalker and E. Charles Brummer (eds.) Legume and crop Genomics. AOCS Press, Champaign, Illinois.

Panthee, D. R., V. R. Pantalone, D. R. West, A. M. Saxton, and C. E. Sams.2005. Quantitative Trait Loci for Seed Protein and Oil Concentration, and Seed Size in Soybean. Crop Sci. 45:2015-2022.

Panthee D.R, V. R. Pantalone, A. M. Saxton. 2006. Modifier QTL for fatty acid composition in soybean oil. Euphytica 152:67-73.

SAS Institute Inc. 2002. SAS Guide for Personal Computers, 9th edn. SAS Institute, Cary, NC.

Sneller, C. H., 1994 Pedigree analysis of elite soybean lines. Crop Sci. 34:1515-1522.

Song, Q. J., L. F. Marek, R. C. Shoemaker, K. G. Lark, V. C. Concibido, X. Delannay, J. E. Specht, and P. B. Cregan, 2004. A new integrated genetic linkage map of the soybean. Theor. Appl. Genet. 109: 122-128.

Soybase. 2008. A genome database for *Glycine*. www.soybase.org; verified April 2, 2008.

Soystats. [online] Available at www.soystats.com (verified 14 April 2008).

Specht, J.E., D.J. Hume, and S.V. Kumudini. 1999. Soybean yield potential- a genetic and physiological perspective. Crop Sci. 39: 1560–1570.

Specht, J. E., K. Chase, M. Macrander, G. L. Graef, J. Chung, J. P. Markwell, M. Germann, J. H. Orf, and K. G. Lark, 2001. Soybean response to water. A QTL analysis of drought tolerance. Crop Sci. 41: 493-509.

Wilcox J.R. 1998. Increasing seed protein in soybean with eight cycles of recurrent selection. Crop Sci.38: 1536-1540.

Wilcox, James R. 2004. World Distribution and Trade of Soybean. *In* H.R Boerma and J. E. Specht (eds.) Soybeans: Improvement, Production, and Uses, 3rd ed. Agron. Monogr. 16. ASA, CSSA, SSSA, Madison, WI.

Wilcox, J. R. Sixty Years of Improvement in Publicly Developed Elite Soybean Lines. 2001. Crop Sci. 41: 1711-1716.

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

Part 3

Detection and Validation of Seed Quality QTL in Soybean

Abstract

Soybean [*Glycine max* (L.) Merrill] is an important commercial crop. Soybean seed quality traits: protein, oil and seed weight determine soybean end use and value. Soybean meal (SBM) is an important source of protein for both human and animal consumption. In 2007, SBM accounted for 69% of the world's fodder consumption. Soybean oil is an important by-product of SBM production which has important dietary and industrial applications. Soybean seed size is especially important in edible soybean markets and it determines end product; for example, large seeded soybeans are preferred for tofu and soymilk.

Soybean seed quality traits are quantitative traits governed by many genes. Use of molecular markers to select for genomic regions governing quantitative traits (QTL) has reduced the time required to produce improved cultivars. The purpose of this study was to validate previously reported seed quality QTL, and submit a manuscript to the Soybean Genetics Committee for review so that confirmed QTL may be named and published for breeder use.

Near infra red spectroscopy (NIRS) was used to predict levels of seed protein and oil and, a 100 seed sample was used to estimate seed weight. Seventy SSR markers placed on linkage groups (LGs) B2, C2, D1a, D1b, D2, F, G, I, H, K, and M were analyzed by single factor ANOVA for possible association with least square means for seed protein, oil and seed weight. Markers were selected based on being in intervals or flanking intervals previously reported to be QTL.

Seed Quality QTL were confirmed as follows: Seed protein [LG C2 (Satt557, Satt460, Satt079 and Satt307), LG F (Satt335, Satt114 and Satt522), LG K (Satt102 and Satt555), LG M (Satt540)], seed oil [LG C2 (Satt307 and Satt202), LG D1a (Satt436), LG D2 (Satt372), LG L (Satt076, Satt166, Satt527 and Satt561) and LG M (Satt590)] and seed weight [LG C2 (Satt557 and Satt079), LG D1a (Satt077 and Satt436), LG F (Satt114 and Satt335), LG K (Satt555) and LG L (Satt076, Satt166, Satt527 and Satt561)]. New QTL were detected on LGs B2, C2, D1a, D1b, D2, F, G, I, K L and M.

Introduction

Soybean [*Glycine max* (L.) Merrill] seed protein, oil and seed weight are important commercial seed quality traits which determine end use and market price. To increase soybean market value, plant breeders have focused breeding efforts on the improvement of these traits to meet market demands (Wilson, 2004).

Soybean seed quality traits are quantitative traits whose expression is governed by many genes and are subject to genotype by environment interaction, which makes selection for those traits difficult. In addition, a negative correlation between protein and oil is well documented in soybean (Burton, 1987), which has made it very difficult to simultaneously increase both traits. Despite these difficulties, plant breeders have achieved much success in improving seed quality traits. Using recurrent selection, Wilcox (1998) successfully selected for higher seed protein content and increased seed protein from 438 to 484 g kg^{-1} in a population derived from two F_2 populations and a high protein cultivar blend. There was a marked decrease in seed oil concentration with every selection cycle, with oil concentration in the population dropping by 23 g kg^{-1}. After cycle five (C5) there was little gain in protein content, which suggested that the population was fixed for favorable alleles.

Greater breeding success can be achieved when chromosomal regions, referred to as quantitative trait loci (QTL), are identified and mapped. Molecular DNA markers are differences in genetic sequences that can be used to differentiate individuals. Examples of molecular DNA markers are: Restriction Fragment Length Polymorphisms (RFLPs), Amplified Fragment Length Polymorphisms (AFLPs), Simple Sequence Repeats (SSRs)

and Single Nucleotide Polymorphisms (SNPs). DNA markers can be used to identify the location of a gene or QTL. Identification of QTL is an essential step in marker assisted selection programs. QTL detection studies must be accurate to avoid the incidence of false positives and false negatives so that the technology continues to support breeding programs (Bernardo, 2002).

Some QTL confirmation studies have been conducted with varying results, which further reiterates the importance of a verification step to support the use of QTL in breeding programs. Sebolt et al. (2000) conducted a confirmation study of QTL controlling protein from a *G. max* x *G. soja* population previously mapped by Diers et al., 1992. In Diers' study QTL controlling protein and oil were mapped on linkage groups (LGs) E and I. A BC_3F_4 line derived from a F_2 (A81-356022 x PI468916) x A81-356022 (recurrent parent) was crossed to two cultivars 'Parker' and 'Kenwood' and the experimental line C1914 to test for the protein QTL. The BC_3F_4 line was selected because it was homozygous for the *G. soja* region on LG I where the protein QTL mapped. Parker, Kenwood and C1419 were used as representatives of cultivars in commercial production because they are high yielding and have average protein content. In the Parker and Kenwood populations, the protein QTL on LG I from PI468916 was associated with increased protein content and a yield penalty. A similar association was not made in the C1419 population and the authors speculate that it contains a gene allelic to that of *Glycine soja* on LG I but does not contain the yield reducing allele from PI 468916. The QTL previously detected on LG E was not significantly associated with protein in this study (Sebolt et al., 2000).

Panthee et al. (2005) conducted a study on QTL for seed size, seed protein and oil concentration in an N87-984-16 by TN93-99 population. They reported seed quality QTL on LGs D1a, D1b, D2, G, H and O. This study highlights the occurrence of environmentally sensitive QTL that are only expressed in certain conditions. These QTL may be useful when selecting for a specific environment, but may be difficult to adapt when selecting for diverse environments (Panthee et al., 2005).

In an Essex x 'Williams' RIL population, Hyten et al. (2004) detected QTL for seed size, oil content and protein content. They were able to detect 6 oil QTL, 4 protein QTL and 7 seed size QTL. Essex and Williams belong to the Southern and Northern germplasm pools respectively, and are the parents of the prominent line 'A3127'. A3127 has been widely used in the pedigrees of many elite lines both in the Northern, and the Southern regions of the United States (Sneller, 1994). Therefore QTL confirmed in our population which is derived from an Essex x Williams 82 cross may be useful to many plant breeders because Williams 82 is a near isogenic line (NIL) of Williams.

The purpose of this study is to: 1) confirm previously reported quantitative trait loci for seed quality in soybean; and 2) submit confirmed QTL as cqQTL to SoyBase (www.soybase.org).

Materials and Methods

Essex x Williams 82 Population Development

The initial crosses for the Essex x Williams 82 population were made at the University of Tennessee Plant Science Farm (KPSF) of the East Tennessee Research and

Education Center (ETREC) in Knoxville, TN in the summer of 2002. The population was advanced from the F_2 to the F_6 generation through single seed descent (Brim, 1966)

Field Experiments

In 2006, 146 recombinant inbred lines (RILs) and the two parents, Essex and Williams 82 were planted in four reps of single plant hill plots (sown 3 beans/hill then thinned to one plant per plot about 14 days after emergence) in a randomized complete block Honeycomb design (Fasoulas and Fasoula, 1995). In addition, 129 genotypes were planted in 3.1m rows sown at 8 beans/0.31m for seed increase for the 2007 experiments. Lines which did not yield enough seed were further increased in winter nursery to ensure that there was sufficient seed for the 2007 experiments.

Based on maturity data collected in 2006, the population was divided by days to maturity into three population subsets: early (22 genotypes), mid (96 genotypes) and late (23 genotypes) separated from each other by ten days in maturity. Five genotypes were dropped from the study because of inadequate seed amounts. In 2007, each population subset which included the two parents was planted in two 6.1m row plots in a randomized complete block design replicated three times at the East Tennessee Research and Education Center (ETREC), the Highland Rim Research and Education Center (HRREC) and the West Tennessee Research and Education Center (WTREC). Checks in the population subsets were assigned by maturity group as follows: early (Macon and LD00-3309), mid (5002T) and late (5002T and 5601T).

Phenotypic Traits

Twenty five grams of soybean seeds were ground in a water-cooled Knifetec 1095 Sample Mill (FOSS Tecator, S-26321, Hogana, Sweden) for 20 seconds. This setting produced soybean flour with a uniform particle size. The near infra red spectroscopy (NIRS) instrument (NIRS 6500, FOSS North America) was warmed up for 2 h after turning on the lamp and auto diagnostics were run. Diagnostics tests ensured the instrument passed three different tests for instrument response, wavelength accuracy, and NIRS repeatability. A room dehumidifier was used throughout the analysis, setting the humidity to 40%. Room temperature was approximately 20ºC. Approximately 15g of the ground soybean samples were scanned using ISIScan software version 2.80. Samples were analyzed according to their replication in the field tests and the instrument was not turned off between reps. Diagnostics were performed every day until all samples were scanned. Sample scanning produced the predicted levels of protein and oil used in data analyses. Seed weight was determined by weighing 100 seeds from each plot.

DNA Extraction and Polymerase Chain Reaction

In 2006, four plants each of 146 RILs and the two parents were grown in the growth chamber for leaf collection for DNA isolation. Leaves were harvested from all four plants at V4 stage. In addition, five leaves per genotype were randomly sampled from the first replication of plots in each test in ETREC in 2007 to ensure that we had enough leaves from which to extract DNA. All leaves were flash frozen in liquid nitrogen and then stored at -80 ºC until DNA was extracted.

DNA was extracted from the leaves of RILs and parents utilizing the Qiagen Plant DNeasy Extraction Kit (Qiagen, Valencia, CA). Polymerase chain reaction (PCR) consisted of 6.9 µL of ddH_2O, 4 µL Hotmaster TAQ (5Prime Gaithersburg, MD), 0.55 µL of 5 µM WellRED labeled forward primer (Proligo, Boulder, CO) and 0.55 µL of 5 µM reverse primer (Sigma-Genosys, Woodlands, TX), and 3 µL of 15 ng/µL template DNA. PCR reactions were performed in a 96-well MBS Hybaid thermocycler (Hybaid, Franklin, MA). PCR conditions consisted of initial denaturation at 95ºC for 2 min followed by 35 cycles of a) denaturation at 92ºC for 1 min, b) 47ºC for annealing for 1 min, c) 68ºC for 5 min for extension and, d) a final cycle at 72ºC for extension for 5 min. Parents were screened with a total of 203 $(ATT)_n$ type simple sequence repeat (SSR) genetic markers developed by Cregan et al. (1999) with the goal to use markers to confirm seed quality trait QTL. Markers were selected based on their proximity to QTL intervals previously reported by Hyten et al. (2004a, 2004b). One hundred and seventeen markers were found to be polymorphic. Seventy of the 117 polymorphic SSR markers which amplified in the progeny were used to genotype RILs for QTL analysis. Sequence information for the SSRs is publicly available from Soybase (www.soybase.org; verified April 3, 2008.)

DNA Gel Electrophoresis

A Beckman-Coulter CEQ 8800 Genome Analysis System (Beckman-Coulter, Fullerton, CA) was used to separate PCR products by gel capillary electrophoresis. Beckman Coulter fragment analysis software was used to estimate PCR amplicon base pair lengths for detectable polymorphisms. Markers were considered to be polymorphic

when the parental amplicon base pair lengths differed by ≥3 nucleotides. Polymorphic primers detected in the parents were used to genotype RILs. RILs were scored (1 = Essex allele, 2 = Heterozygote, and 3 = Williams 82 allele) for all markers including pubescence color and flower color. Heterozygote scores were dropped from the data set so as to only report potential additive genetic effects.

Data Analysis

The early, mid, and late population subsets were combined across the three environments (ETREC, HRREC and WTREC) to create three populations named set-1, set-2 and set-3 respectively. It was hypothesized that by separating the populations by maturity group we might be able to see QTL differences within a more narrow range of maturity. Sets 1, 2 and 3 were also combined to form the total population which could enable us to further detect QTL with greater power.

Phenotypic data was analyzed using the MIXED procedure of SAS (SAS Institute Inc., 2002) software to determine whether there were significant genotypic differences among lines within sets and within the total population. Environments and replication were considered as two blocking factors in the model.

QTL were analyzed via single factor ANOVA in SAS using the GLM procedure (SAS Institute Inc., 2002) procedure to determine if there were any associations between SSR markers and phenotypic least square means for the traits of interest. MAPMAKER /EXP 3.0 (Lander et al. 1987; Lincoln et al. 1992) was used to estimate genetic distances between SSR markers. A minimum likelihood of odds (LOD) ≥3.0 and a maximum distance ≤50cM were used to test linkage among markers. Marker order was similar to

the soybean composite integrated map (Song et al. 2004) a finding that was also supported in the Essex x Williams population by Hyten et al. (2004a). The soybean composite integrated map was used to determine marker distances because it is based on five mapping populations and marker distances are more accurate.

Heritability for each trait was calculated on an entry mean basis for three environments and three replications according to Nyquist (1991) whose formula was presented by Panthee et al. (2005).

REML estimation in PROC MIXED was used to determine the variance components for heritability calculations. PROC CORR (SAS 2002) was used to determine phenotypic correlations between traits.

Results and Discussion

A total of seventy SSR markers placed on LGs B2, C2, D1a, D1b, D2, F, G, H, I, K, L and M were analyzed by single factor ANOVA for their possible associations with least square means for seed protein, oil and seed weight in the confirmation population.

There were significant genotypic differences for seed protein among the RILs in the confirmation population, in all population sets and in the total population. Seed protein concentration ranged from 314 to 424 g kg^{-1} , with a mean of 379 g kg^{-1} in the total RIL population (Table 3.1). Parental means were similar so the wide range of RIL protein concentration observed was attributed to transgressive segregation. There was significant negative correlation between protein and oil ($r = -0.29$, $p < 0.001$) and a positive correlation between seed protein and seed weight ($r=0.31$, $p < 0.001$) (Table 3.2).

Table 3.1 Descriptive statistics for seed quality traits in an F_6 derived soybean RIL population from an Essex x Williams 82 cross grown in three environments Tennessee

Trait	Min	Max	Mean	Std. Dev	Essex	Williams 82	h^2
Protein[a]	338	408	379	9.8	385	390	0.92
Oil[a]	184	239	196	6.5	189	200	0.91
Seed weight[b]	110	169	141	11.0	137	155	0.92

[a] Protein and oil concentration in g kg^{-}
[b] Seed weight in mg $seed^{-1}$

Table 3.2 Phenotypic correlation coefficients among seed quality traits in an F_6 derived soybean RIL population from an Essex Williams 82 grown in three environments in Tennessee

	Protein	Oil	Seed weight
Protein		-0.2906***	0.3077***
Oil	-0.2906***		0.1737*
Seed weight	0.3077***	0.1737*	

*, **, ***, ****Significant at P=0.05, P=0.01, P=0.001and P=0.0001; ns, non-significant

Correlations between seed protein and oil and, seed protein seed weight are consistent with the literature (Csanadi et al., 2001; Panthee et al., 2005).

Protein was a highly heritable trait in this population (0.95) which suggests that most of the phenotypic variation observed was due to genetic effects.

Hyten et al. (2004a) detected four protein QTL on LGs C2, F, K and M in an Essex x Williams population. QTL on LGs C2, F and K were significant in all combined environments while the QTL on LG M was only detected in one environment. A seed protein QTL on LG C2 was detected in the interval Satt277-Satt202 in the original mapping population. In the confirmation study, P3 (set-1 and the total population), P4 (set-1, set-2 and the total population), P5 (set-1) and P6 (total population) were confirmed in this interval (Table 3.3).

The QTL P5 (total population) and P6 (set-1) were also detected on C2 but were not confirmed because their p-values (0.042 and 0.015, respectively) exceeded the confirmation threshold of $p < 0.01$ established by the Soybean Genetics Committee. P21–P23 were confirmed in the interval Satt114-Satt335 on LG F in set-2 and in the total population. A protein QTL on LG K in the interval Satt539-Satt102 was confirmed by P25 (set-1) and P26 (set-2). P36 (set-1) confirmed the protein QTL on LG M detected in the interval Satt540-Satt463.

Other protein QTL not reported by Hyten et al. (2004a) were detected on LGs B2, C2, D1a, D1b, F, G, K and L. P2 (set-2 and the total population) on LG C2 is approximately 50cM upstream from the closest protein QTL reported in Soybase, so it is likely a new QTL. P8 (set-2 and set-3) and P9 (set-3) on LG D1a are at least 10cM downstream from the closest QTL reported in Soybase. P8 and P9 are less than one 1cM

Table 3.3 QTL, map position and additive genetic effects for seed protein concentration in F_6 derived early, medium, and late maturing subpopulations and the total population from an Essex x Williams 82 cross grown in three environments in Tennessee

				Set-1			Set-2			Set-3			Total Population		
QTL[a]	**Marker[b]**	**LG[c]**	**Pos[d]**	**Effect[e]**	**P**	**R^2**	**Effect[e]**	**P**	**R^2**	**Effect[e]**	**P**	**R^2**	**Effect[e]**	**P**	**R^2**
P1	Satt304	B2	65.56	8	0.0276	0.08	-	ns	-	-	ns	-	-	ns	-
P2	Satt291	C2	45.78	-	ns	-	6	0.0008	0.05	-	ns	-	3	0.0395	0.03
P3	Satt557	C2	112.19	-17	0.0001	0.51	-	ns	-	-	ns	-	-15	0.0001	0.22
P4	Satt460	C2	117.77	-15	0.0001	0.35	-8	0.0027	0.03	-	ns	-	-10	0.0001	0.15
P5	Satt079	C2	117.87	-13	0.0001	0.24	-	ns	-	-	ns	-	-4	0.042	0.03
P6	Satt307	C2	121.27	-8	0.015	0.09	-	ns	-	-	ns	-	-6	0.0022	0.07
P7	Satt202	C2	126.23	-	ns	-	-	ns	-	-	ns	-	-5	0.0074	0.06
P8	Satt129	D1a	11.2	-	ns	-	-5	0.0061	0.04	9	0.0323	0.09	-	ns	-
P9	Satt147	D1a	11.99	-	ns	-	-	ns	-	10	0.0078	0.12	-	ns	-
P10	Satt077	D1a	43.39	-13	0.0001	0.24	-	ns	-	-	ns	-	-3	0.06	0.03
P11	Satt436	D1a	50.19	-14	0.0001	0.23	-4	0.0152	0.02	-	ns	-	-	ns	-
P12	Satt203	D1a	61.89	-18	0.0001	0.46	-	ns	-	-	ns	-	-4	0.0143	0.05
P13	Satt179	D1a	64.69	-5	0.0481	0.06	-	ns	-	-	ns	-	-	ns	-
P14	Satt266	D1b	59.61	-	ns	-	-5	0.01	0.03	-	ns	-	-6	0.0055	0.07
P15	Satt412	D1b	72.57	-	ns	-	5	0.034	0.03	-	ns	-	-	ns	-
P16	Satt014	D2	29.56	7	0.0253	0.08	-	ns	-	-	ns	-	-	ns	-
P17	Satt649	F	5.361	-	ns	-	-4	0.02	0.02	-	ns	-	-	ns	-
P18	Satt348	F	15.29	-	ns	-	-5	0.0141	0.02	-	ns	-	-3	0.0414	0.04
P19	Satt252	F	16.08	-	ns	-	-5	0.0053	0.03	-	ns	-	-4	0.0144	0.05
P20	Satt114	F	63.69	-	ns	-	6	0.0018	0.05	-	ns	-	6	0.0031	0.09
P21	Satt335	F	77.7	-	ns	-	5	0.0013	0.04	-	ns	-	5	0.0017	0.08
P22	Satt144	F	102.08	-	ns	-	7	0.0005	0.05	-	ns	-	6	0.0045	0.07
P23	Satt522	F	119.19	-	ns	-	7	0.0001	0.07	-	ns	-	4	0.017	0.06

Table 3.3 Continued

					Set-1			Set-2			Set-3			Total Population	
QTL[a]	**Marker[b]**	**LG[c]**	**Pos[d]**	**Effect[e]**	**P**	**R^2**	**Effect[e]**	**P**	**R^2**	**Effect[e]**	**P**	**R^2**	**Effect[e]**	**P**	**R^2**
P24	Satt191	G	96.57	9	0.0499	0.13	-8	0.0149	0.07	-	ns	-	-	ns	-
P25	Satt102	K	30.28	15	0.0004	0.31	-	ns	-	-	ns	-	-	ns	-
P26	Satt555	K	42.71	-	ns	-	-5	0.003	0.04	-	ns	-	-	ns	-
P27	Satt273	K	56.62	-	ns		-4	0.0475	0.02	-	ns	-	-	ns	-
P28	Satt725	K	56.85	9	0.0051	0.13	-5	0.0133	0.02	-	ns	-	-	ns	-
P29	Satt475	K	78.68	-	ns	-	-4	0.0079	0.03	-	ns	-	-4	0.0345	0.04
P30	Satt481	L	54.57	-10	0.0006	0.19	-	ns	-	-	ns	-	-4	0.0092	0.06
P31	Satt156	L	56.14	-9	0.0355	0.11	-	ns	-	-	ns	-	-5	0.016	0.06
P32	Satt076	L	61.35	-16	0.0001	0.44	-5	0.0029	0.04	-9	0.0299	0.08	-7	0.0001	0.16
P33	Satt166	L	66.55	-15	0.0001	0.41	-	ns	-	-	ns	-	-6	0.0003	0.11
P34	Satt527	L	70.36	-19	0.0001	0.54	-5	0.0024	0.04	-	ns	-	-7	0.0001	0.15
P35	Satt561	L	71.44	-13	0.0001	0.3	-5	0.0133	0.03	-	ns	-	-8	0.0002	0.12
P36	Satt540	M	35.85	9	0.0028	0.14	-	ns	-	-	ns	-	-	ns	-

[a] P1 to P36 denote thirty six seed protein QTL detected
[b] SSR marker associated with the QTL by SF ANOVA
[c] Linkage group based on the integrated soybean genetic linkage map
[d] Marker position on linkage group based on the integrated soybean genetic linkage map
[e] Average change (g kg^{-1}) with respect to Essex allele

apart, so this is probably a single QTL. P10 (set-1 and the total population) and P11 (set-1 and set-2) have previously been reported (Brummer et al., 1997).

P12 (set-1 and the total population) and P13 (set-1) are approximately 4cM apart and at least 12cM downstream from the closest QTL reported in Soybase and likely constitute a single new QTL on D1a. P14 (set-2 and the total population) and P15 (set-2) detected on LG D1b are at least 20cM downstream from the closest protein QTL reported in Soybase so they may be new QTL. P16 (set-2) on LG D2 may be a new QTL as there no protein QTL reported in Soybase on this linkage group.

There were significant genotypic differences for seed oil among the RILs in the confirmation population, in all population sets and in the total population ($p < 0.0001$). Seed oil concentration ranged from 184 to 239 g kg^{-1}, with a mean of 141 g kg^{-1} in the total RIL population (Table 3.1). There was significant negative correlation between seed oil and protein($r = -0.29$, $p < 0.001$) and a positive correlation between seed oil and seed weight ($r=0.17$, $p < 0.05$) (Table 3.2). Correlations between seed oil and protein and seed weight are consistent with the literature (Burton, 1998; Wilson 2004). The high heritability estimate (0.91) for seed oil in this population suggests that most of the phenotypic variation observed among RILs is due to genetic effects.

Hyten et al. (2004a) mapped six oil QTL on LGs C2, D1a, D2, L and M. The oil QTL on LG C2 mapped in the interval Satt277-Satt460 was confirmed by O7 (set-2) and O8 (set-2) (Table 3.4). O10 (set-3) confirmed the oil QTL detected on LG D1a in the interval Satt184-Satt 373. O9 (set-3) on LG D1a was detected but not confirmed ($0.01 < p < 0.05$). O14 (set-2) confirmed the oil QTL previously detected on LG D2 in the interval Satt458-Satt154. Two oil QTL were detected on LG L in the interval Satt166-

Table 3.4 QTL, map position and additive genetic effects for seed oil concentration in F_6 derived early, medium, and late maturing subpopulations and the total population from an Essex x Williams 82 grown in three environments in Tennessee

				Set-1			Set-2			Set-3			Total Population		
QTL[a]	**Marker[b]**	**LG[c]**	**Pos[d]**	**Effect[e]**	**P**	**R^2**	**Effect[e]**	**P**	**R^2**	**Effect[e]**	**P**	**R^2**	**Effect[e]**	**P**	**R^2**
O1	Satt304	B2	65.56	3	0.0475	0.07	-	ns	-	ns	ns	-	-	ns	-
O2	Satt070	B2	72.8	-	ns	-	-3	0.0003	0.06	ns	ns	-	-	ns	-
O3	Satt474	B2	75.34	-	ns	-	-2	0.0089	0.03	-	-	-	-	-	-
O4	Satt534	B2	87.59	-	ns	-	3	0.0366	0.02	-	-	-	-	-	-
O5	Satt291	C2	96.57	-	ns	-	-4	0.0054	0.03	-	-	-	-	-	-
O6	Satt079	C2	117.87	-	ns	-	3	0.0359	0.02	-	-	-	-	-	-
O7	Satt307	C2	121.27	-	ns	-	4	0.0013	0.04	-	-	-	-	-	-
O8	Satt202	C2	126.24	-	ns	-	4	0.0045	0.03	-	-	-	-	-	-
O9	Satt077	D1a	43.39	-	ns	-	-2	0.034	0.02	-3	0.0213	0.12	-3	0.0181	0.05
O10	Satt436	D1a	50.19	-	ns	-	-	-	-	-6	0.0037	0.16	-3	0.0391	0.04
O11	Satt412	D1b	72.57	-	ns	-	2	0.0144	0.04	6	0.0448	0.17	-	-	-
O12	Satt546	D1b	87.2	-	ns	-	-3	0.0277	0.02	-5	0.0126	0.13	-	-	-
O13	Satt271	D1b	137.06	-	ns	-	-	-	-	5	0.0227	0.10	-	-	-
O14	Satt372	D2	39.35	-	ns	-	-4	0.0057	0.03	-	-	-	-3	0.0127	0.06
O15	Satt461	D2	80.19	-	ns	-	-	-	-	-6	0.0037	0.13	-	-	-
O16	Satt649	F	5.361	6	0.0105	0.16	-	-	-	-	-	-	-	-	-
O17	Satt269	F	11.37	7	0.0001	0.33	-	-	-	-	-	-	4	0.0177	0.06
O18	Satt252	F	16.08	-	ns	-	4	0.004	0.04	-	-	-	4	0.0064	0.07
O19	Satt149	F	18.13	-	ns	-	4	0.0134	0.03	-	-	-	2	0.0324	0.04
O20	Satt335	F	77.7	4	0.0337	0.08	-4	0.0048	0.03	-	-	-	-	-	-
O21	Satt144	F	102.08	7	0.0001	0.27	-	-	-	-	-	-	-	-	-
O22	Satt522	F	119.19	6	0.0033	0.15	-		-	-	-	-	-	-	-

Table 3.4 Continued

				Set-1			Set-2			Set-3			Total Population		
QTL[a]	**Marker[b]**	**LG[c]**	**POS[d]**	**Effect[e]**	**P**	**R^2**	**Effect[e]**	**P**	**R^2**	**Effect[e]**	**P**	**R^2**	**Effect[e]**	**P**	**R^2**
O23	Satt324	G	33.26	-	ns	-	-4	0.0071	0.04	-	-	-	-2	0.0415	0.04
O24	Satt270	I	50.11	-	ns	-	-	-	-	-5	0.0096	0.13	-	-	-
O25	Satt240	K	52.88	-	ns	-	2	0.0391	0.02	-	-	-	-	-	-
O26	Satt273	K	56.62	-5	0.0052	0.12	-	-	-	-	-	-	-	-	-
O27	Satt725	K	56.85	-6	0.0006	0.19	-	-	-	-	-	-	-	-	-
O28	Satt475	K	78.68	-6	0.0021	0.16	-	-	-	-	-	-	-	-	-
O29	Satt260	K	80.12	-4	0.0357	0.09	-3	0.0001	0.06	-	-	-	-2	0.023	0.05
O30	Satt143	L	30.19	-	ns	-	-2	0.0414	0.02	-	-	-	-	-	-
O31	Satt481	L	54.57	-	ns	-	-4	0.0001	0.11	-	-	-	-3	0.0054	0.07
O32	Satt076	L	61.35	-	ns	-	-3	0.0041	0.03	-	-	-	-	-	-
O33	Satt166	L	66.513	-	ns	-	-5	0.0003	0.06	-	-	-	-4	0.0027	0.08
O34	Satt527	L	70.36	-	ns	-	-4	0.0001	0.09	-	-	-	-3	0.0027	0.08
O35	Satt561	L	71.44	-	ns	-	-4	0.005	0.04	-	-	-	-3	0.0239	0.05
O36	Satt229	L	93.89	-	ns	-	-3	0.0219	0.02	-	-	-	-3	0.0348	0.04
O37	Satt590	M	7.841	-5	0.0044	0.16	-3	0.0207	0.02	-6	0.0012	0.22	-4	0.0029	0.08

[a] 01 to 037 denote thirty seven seed oil QTL detected
[b] SSR marker associated with the QTL by SF ANOVA
[c] Linkage group based on the integrated soybean genetic linkage map
[d] Marker position on linkage group based on the integrated soybean genetic linkage map
[e] Average change (g kg^{-1}) with respect to Essex allele

Dt1 and Satt229-Satt373 (Hyten et al., 2004a). O32 (set-2), O33 (total population), O34 (set-2 and the total population), O35 (set-2) were confirmed in the interval Satt166-Dt1. O36 was detected but not confirmed p-value ($0.01 < p < 0.05$). In the confirmation population, O36 was not linked to the interval described by Hyten et al. (2004) perhaps because there was a scarcity of polymorphic markers in this region of LG L. QTL previously detected on LG M was confirmed by O37 (set-1, set-3 and the total population).

Other oil QTL previously not reported by Hyten et al. (2004a) were detected on LGs B2, C2, D1b, F, G, I and K in this study. O1 (set-1) and O2-O4 (set-2) detected on LG B2 have previously been reported in Soybase (2008). O11 and O12 on LG D1b detected in sets 2 and 3 are approximately 5cM apart and probably constitute a single QTL. This interval is approximately 35cM upstream from the closest oil QTL reported in Soybase so it is a likely new QTL. O13 (set-2) is about 50cM downstream from the O11-O12 interval, and at least 15cM downstream from the closest oil QTL in reported in Soybase so it is likely to be another new oil QTL on this linkage group. O15 (set-3), approximately 40cM downstream from O14 on LG D2 has been previously reported in Soybase (2008).

O16 (set-1), O18 (set-1 and the total population), O19 and O20 (set-3 and the total population) were detected on LG F. The interval O17-O19 is about 13cM long and the largest approximate distance between any two markers is 5cM. O17-O19 therefore likely constitutes a single new QTL, as no other QTL have been reported on LG F close to this interval. O20 (set1 and set-2) on F is about 25cM downstream from O17-O19 and has previously been reported in Soybase (2008). O21 (set-1) was detected 11cM upstream

from the closest oil QTL on LG F reported in Soybase and is probably another a new QTL. O22 on LG F has previously been reported in Soybase (2008).

O23 (set-2 and the total population) on LG G is about 30cM from the closest QTL reported in Soybase and may be a new QTL. O24 detected on LG I has been previously detected by other researchers as an important QTL governing seed oil and protein (Diers et al., 1992; Sebolt et al., 2000; Chung et al., 2003). Nichols et al. (2006) have confirmed QTL for seed yield (cqSd yld -001), maturity (cqPod mat-001) and seed weight (cqSd wt-003) in this region in a population of near isogenic lines derived from a *G. max* x *G. soja* cross. O24 was not significantly associated with protein in this study. The detection of this QTL in our population may indicate that it is not unique to *G. soja* alleles or that it may be a modifier in G. max. O25 (set-2), O26 (set-1) and O27 (set-1) on LG K were not previously reported in Soybase. The O25-O27 interval is about 4cM wide so O25, O26, and O27 probably constitute a single new QTL. The interval O28-O29 (approx. 2cM wide) is 10cM away from the closest QTL in reported in Soybase and may be another new QTL on this linkage group.

There were significant genotypic differences for seed weight among the RILs in the confirmation population within all population sets, and within the total population ($p < 0.0001$). Seed weight concentration ranged from 110 to 169 mg seed $^{-1}$, with a mean of 196 mg 141 mg seed $^{-1}$ in the total RIL population (Table 3.1). Seed weight was positively correlated with protein ($r= 0.31$, $p< 0.001$) and oil ($r = 0.17$, $p < 0.05$) (Table 3.2). Correlations between seed weight and, protein and oil are consistent with the literature (Wilson, 2004). The high heritability estimate (0.92) for seed weight implies that most of the phenotypic variation observed is due to genetic effects.

Seven seed weight QTL were mapped by Hyten et al. (2004a) on LGs C2, D1a, F, G, I, K and L. A seed weight QTL on LG C2 in the interval Satt277-Satt460 was confirmed by SW3 (set-1 and set-3), SW4 (set-3) and SW6 (set-2) (Table 3.5). SW5 (set-2) was detected but not confirmed in this interval ($0.01<p<0.05$). The seed weight QTL in the interval Satt179-Satt071on LG D1a was confirmed by SW9–SW10 (set-1, set-2 and the total population). SW17-SW18 (set-1, set-2 and the total population) confirmed the seed weight QTL on LG F in the interval Satt114-Satt335.

Seed weight QTL on LG K were previously mapped in the interval Satt518-Satt273. SW23 (set-2) was confirmed in this interval. Other QTL within this interval SW23 (total population) and SW24 (set-2) were detected but not confirmed ($0.01 <p <0.05$). Seed weight QTL were reported on LG L in the interval Satt156-Dt1. SW28, SW30-SW31 (set-1, set-2 and the total population) and SW29 (set-2 and the total population) were confirmed in this interval. SW27 (set-2) and SW29 (set-1) were detected but not confirmed ($0.01< p <0.05$).

Other seed weight QTL previously not reported by Hyten et al. (2004a) were detected on LGs B2, C2, D1b, F, G, H and K in this study. Seed size QTL SW1 (set-1) and SW2 (set-2) detected on LG B2 have been previously reported in Soybase. SW13 (set-2 and set-3) on LG D1b has not been previously detected and is probably a new QTL detected in this study. SW14 (set-2) and SW15 (set 2 and set-3) on LG D2 have been reported in Soybase. SW16 detected on LG F is approximately 25 cM upstream from the closest seed weight QTL reported in Soybase and is likely a new QTL. SW20 detected on LG G has previously been reported in Soybase. SW25 (set-2) and SW26 (set-2 and the

Table 3.5 QTL, map position and additive genetic effects for seed weight in F_6 derived early, mid, late maturing subpopulations and the total population from an Essex x Williams 82 grown in three environments in Tennessee

				Set-1			Set-2			Set-3			Total population		
QTL[a]	**Marker[b]**	**LG[c]**	**Pos[d]**	**Effect[e]**	**P**	**R^2**	**Effect[e]**	**P**	**R^2**	**Effect[e]**	**P**	**R^2**	**Effect[e]**	**P**	**R^2**
SW1	Satt304	B2	65.56	10	0.033	0.08	-	-	-	-	-	-	-	-	-
SW2	Satt474	B2	75.34	-	-	-	-3	0.0493	0.02	-	-	-	-	-	-
SW3	Satt557	C2	112.19	-17	0.0041	0.2	-	-	-	-24	0.0023	0.22	-	-	-
SW4	Satt079	C2	117.87	-	-	-	-	-	-	-12	0.009	0.11	-	-	-
SW5	Satt307	C2	121.27	-	-	-	-4	0.0248	0.02	-	-	-	-	-	-
SW6	Satt202	C2	126.24	-	-	-	-4	0.0123	0.02	-	-	-	-	-	-
SW7	Satt129	D1a	11.2	-	-	-	-5	0.0042	0.04	-	-	-	-	-	-
SW8	Satt147	D1a	11.99	-	-	-	-4	0.0129	0.03	-	-	-	-	-	-
SW9	Satt077	D1a	43.39	-19	0.0001	0.24	-7	0.0001	0.07	-	-	-	-9	0.0001	0.14
SW10	Satt436	D1a	50.19	-17	0.0005	0.19	-6	0.0004	0.05	-	-	-	-9	0.0001	0.12
SW11	Satt203	D1a	61.89	-19	0.0002	0.23	-7	0.0001	0.08	-	-	-	-8	0.0001	0.14
SW12	Satt179	D1a	64.69	-13	0.0025	0.14	-5	0.0021	0.04	-7	0.0195	0.09	-6	0.0004	0.1
SW13	Satt412	D1b	72.57	-	-	-	4	0.0335	0.03	-14	0.0168	0.23	-	-	-
SW14	Satt014	D2	29.56	-	-	-	-3	0.0494	0.02	-	-	-	-	-	-
SW15	Satt372	D2	39.35	-	-	-	-4	0.0076	0.03	-5	0.0363	0.09	-	-	-
SW16	Satt348	F	15.29	-	-	-	-	-	-	-	-	-	-5	0.0213	0.04
SW17	Satt114	F	63.69	10	0.034	0.09	7	0.0005	0.06	-	-	-	8	0.0005	0.12
SW18	Satt335	F	77.7	11	0.0197	0.09	7	0.0001	0.08	-	-	-	7	0.0002	0.11
SW19	Satt522	F	119.19	-	-	-	5	0.0157	0.03	-9	0.0279	0.11	-	-	-

Table 3.5 Continued

					Set-1			Set-2			Set-3		Total population		
QTL[a]	**Marker[b]**	**LG[c]**	**Pos[d]**	**Effect[e]**	**P**	**R^2**	**Effect[e]**	**P**	**R^2**	**Effect[e]**	**P**	**R^2**	**Effect[e]**	**P**	**R^2**
SW20	Satt191	G	96.57	15	0.0481	0.13	-6	0.0495	0.05	-	-	-	-	-	-
SW21	Satt317	H	89.52	14	0.0391	0.09	-	-	-	-	-	-	-	-	-
SW22	Satt102	K	30.28	13	0.032	0.13	-5	0.0121	0.03	-	-	-	-	-	-
SW23	Satt555	K	42.71	-	-	-	-6	0.0001	0.07	-	-	-	-4	0.0289	0.04
SW24	Satt725	K	56.85	-	-	-	-4	0.0254	0.02	-	-	-	-	-	-
SW25	Satt475	K	78.68	-	-	-	-5	0.0015	0.04	-	-	-	-	-	-
SW26	Satt260	K	80.12	-	-	-	-7	0.0001	0.07	-	-	-	-4	0.0283	0.04
SW27	Satt481	L	54.57	-	-	-	-3	0.0374	0.02	-	-	-	-	-	-
SW28	Satt076	L	61.35	-13	0.0029	0.15	-8	0.0001	0.11	-	-	-	-9	0.0001	0.17
SW29	Satt166	L	66.51	-11	0.0117	0.1	-10	0.0001	0.16	-	-	-	-9	0.0001	0.17
SW30	Satt527	L	70.36	-17	0.0006	0.21	-8	0.0001	0.11	-	-	-	-10	0.0001	0.18
SW31	Satt561	L	71.44	-12	0.0052	0.13	-10	0.0001	0.16	-	-	-	-10	0.0001	0.2
SW32	Satt229	L	93.89	-	-	-	-6	0.0003	0.06	-	-	-	-5	0.0059	0.07
SW33	Satt540	M	35.85	9	0.0351	0.07	-	-	-	-	-	-	-	-	-

[a] SW1 to SW33 denote thirty three seed weight QTL detected

[b] SSR marker associated with the QTL by SF ANOVA

[c] Linkage group based on the integrated soybean genetic linkage map

[d] Marker position on linkage group based on the integrated soybean genetic linkage map

[e] Average change (mg $seed^{-1}$) with respect to Essex allele

total population) lie close to other seed weight QTL previously reported in this region (Mian et al., 1996). SW33 (set-1) detected on LG M has been previously reported in an independent Essex x Williams population (Chapman et al., 2003), but was not confirmed because its p-value (0.0341) exceeds the confirmation threshold of $p < 0.01$.

Conclusion

Several QTL governing seed quality traits have been confirmed in this study. Examples of confirmed QTL are as follows: Protein (P3, P4, P5, P6, and P36), oil (O7, O8, O10, and O14) and seed weight (SW3, SW4, SW5, SW6 and SW18).

There were notable differences in the number of QTL detected for each trait among population sets. The most notable difference being between the number of QTL detected for protein and seed weight between population sets 1 and 3. There were 20 protein and seed weight QTL set-1 in contrast to only 3 in set-3. There were 16 seed weight detected in set-1 and 6 in set-3.

The difference in the number of maturity QTL detected between set-1 and set-3 suggests that maturity might be a strong influence on seed weight and protein expression. Nichols et al. (2006) confirmed QTL conditioning for seed protein, oil and maturity in the same region on LG I. However, as, their study was focused on one region on LG I, more studies, with larger population sizes segregating for maturity could help investigate which loci most affect maturity, seed weight, protein and oil simultaneously.

References

References

Brim, C. A., 1966. A modified pedigree method of selection in soybeans. Crop Sci. 6: 220.

Brummer, E. C., G. L. Graef, J. Orf, J. R. Wilcox, and R. C. Shoemaker, 1997. Mapping QTL for seed protein and oil content in eight soybean populations. Crop Sci. 37: 3 70-378.

Chapman, A., V. R. Pantalone, A. Ustun, F. L. Allen, D. Landau Ellis, R. N. Trigiano, and P. M. Gresshoff, 2003. Quantitative trait loci for agronomic and seed quality traits in an F2 and F4.6 soybean population. Euphytica 129:387-393.

Cregan, P. B., T. Jarvik, A. L. Bush, R. C. Shoemaker, K. G. Lark, A. L. Kahler, N. Kaya, T. T. VanToai, D. G. Lohnes, and J. Chung, 1999. An integrated genetic linkage map of the soybean genome. Crop Sci. 39: 1464—1490.

Csanadi, G., J. Vollmann, G. Stift, and T. Lelley. 2001. Seed Quality QTLs identified in a molecular map of early maturing soybean. Theor. Appl. Genet. 103: 912-919.

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

Fasoula, Vasilia A., Donna K. Harris and H. Roger Boerma. 2004. Validation and designation of quantitative trait loci for seed protein, seed oil, and seed weight from two soybean populations. Crop Sci. 44: 1218-1225.

Fasoulas, A.C and V.A Fasoula 2005. Honey comb selection designs. *In* Jules Janick (ed.) Plant Breeding Reviews. Vol. 13. John Wiley and Sons Inc. New York, USA.

Fehr, W.R., Caviness, C.E., Burmond, D.L., and Pennington, J.S. 1971. Stages of development decriptions for soybeans for soybeans, Glycine max (L.) Merrill. Crop Sci. 11: 929-931.

Hyten, D. L. QTL mapping and identification of GxE interactions of agronomic and seed quality Traits in soybean. 2002.

Hyten, D. L., V. R. Pantalone, C. E. Sams, A. M. Saxton, D. Landau-Ellis, T. R. Stefaniak, and M. E. Schmidt, 2004a. Seed quality QTL in a prominent soybean population. Theor. Appl. Genet. 109: 552—561.

Hyten, D. L., Vincent R. Pantalone, Arnold M. Saxton, Michael E. Schmidt and Carl E. Sams. 2004b. Molecular mapping and identification of soybean fatty acid modifier quantitative trait loci. J. Am. Oil Chem. Soc. 81 (12):1115-1118.

Lee, S. H., M. A. Bailey, M. A. R. Mian, E. R. Shipe, D. A. Ashley, W. A. Parrott, R. S. Hussey, and H. R. Boerma, 1996. Identification of quantitative trait loci for plant height, lodging, and maturity in a soybean population segregating for growth habit. Theor. Appl. Genet. 92: 516-523.

Lincoln, S. E., M. J. Daly, and E. S. Lander, 1993. Constructing Genetic Linkage Maps with MAPMAKER, Version 3. A Tutorial and Reference Manual. Release 3rd ed. Whitehead Institute for Biomedical Research, Cambridge, MA.

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

Mian, M. A. R., M. A. Bailey, J. P. Tamulonis, E. R. Shipe, T. E. Carter Jr, W. A. Parrott, D. A. Ashley, R. S. Hussey, and H. R. Boerma, 1996. Molecular markers associated with seed weight in two soybean populations. Theor. Appl. Genet. 93: 1011-1016.

Nyquist, W. E., 1991. Estimation of heritability and prediction of selection response in plant populations. Crit. Rev. Plant Sci. 10: 235-322.

Orf, J. H., K. Chase, T. Jarvik, L. M. Mansur, P. B. Cregan, F. R. Adler, and K. G. Lark, 1999b. Genetics of soybean agronomic traits. I. Comparison of three related recombinant inbred populations. Crop Sci. 39: 1642-1651.

Pantalone, Vincent R., David R. Walker, Ralph Dewey and Istvan Rajcan. 2004. DNA Marker-Assisted Selection for Improvement of Soybean Oil Concentration and Quality. p. 284-285, 301-303. *In* H. Thomas Stalker and E. Charles Brummer (eds.) Legume and crop Genomics. AOCS Press, Champaign, Illinois.

Panthee, D. R., V. R. Pantalone, D. R. West, A. M. Saxton, and C. E. Sams.2005. Quantitative Trait Loci for Seed Protein and Oil Concentration, and Seed Size in Soybean. Crop Sci. 45: 2015-2022.

Panthee D.R, V. R. Pantalone, A. M. Saxton. 2006. Modifier QTL for fatty acid composition in soybean oil. Euphytica 152: 67-73.

SAS Institute Inc. 2002. SAS Guide for Personal Computers, 9th edn. SAS Institute, Cary, NC.

Sneller, C.H. 1994. Pedigree analysis of elite soybean lines. Crop Sci. 34: 1515-1522.

Song, Q. J., L. F. Marek, R. C. Shoemaker, K. G. Lark, V. C. Concibido, X. Delannay, J. E. Specht, and P. B. Cregan, 2004. A new integrated genetic linkage map of the soybean. Theor. Appl. Genet. 109: 122-128.

Soybase. 2008. A genome database for *Glycine*. www.soybase.org; verified April 2, 2008.

Soystats. [online] Available at www.soystats.com (verified 14 April 2008).

Specht, J.E., D.J. Hume, and S.V. Kumudini. 1999. Soybean yield potential- a genetic and physiological perspective. Crop Sci. 39: 1560–1570.

Specht, J. E., K. Chase, M. Macrander, G. L. Graef, J. Chung, J. P. Markwell, M. Germann, J. H. Orf, and K. G. Lark, 2001. Soybean response to water. A QTL analysis of drought tolerance. Crop Sci. 41: 493-509.

Wilcox J.R. 1998. Increasing seed protein in soybean with eight cycles of recurrent selection. Crop Sci.38: 1536-1540.

Wilcox, James R. 2004. World Distribution and Trade of Soybean. *In* H.R Boerma and J. E. Specht (eds.) Soybeans: Improvement, Production, and Uses, 3rd ed. Agron. Monogr. 16. ASA, CSSA, SSSA, Madison, WI.

Wilcox, J. R. Sixty Years of Improvement in Publicly Developed Elite Soybean Lines. 2001. Crop Sci. 41: 1711-1716.

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

Part 4

Detection and Validation of Soybean Fatty acid Modifier QTL

Abstract

Soybean [*Glycine max* (L.) Merrill] oil primarily contains the following fatty acids: palmitic, stearic, oleic, linoleic, and linolenic. The concentration and characteristics of each fatty acid determine the quality of products produced. For example, linolenic acid is prone to oxidation, which turns the oil product rancid thus reducing its shelf life. Producing soybeans which have specific fatty acid content is importance for edible oil and industrial purposes. Several modifier genes have been found to influence palmitic, stearic, oleic, linoleic and linolenic acid content in soybean. The purpose of this study was to confirm previously reported quantitative trait loci (QTL) for fatty acid modifier genes in an independent 'Essex' x 'Williams 82' population, report novel fatty acid QTL, and submit a manuscript to the Soybean Genetics Committee for review so that confirmed QTL may be named and published in Soybase for breeder use.

Fatty acid analysis by gas chromatography of the methyl esters for 131 F_6 derived recombinant inbred lines (RIL) and the two parents was performed. The RILs averaged 124 g kg^{-1} palmitic acid, 42 g kg^{-1} stearic acid, 245 g kg^{-1} oleic acid, 531 g kg^{-1} linoleic acid and 59 g kg^{-1} linolenic acid. Parents were screened with 203 SSR markers on linkage groups (LGs) B2, C2, D1a, D1b, D2, F, G, H, I, K, L and M, previously linked to fatty acid modifier QTL. Polymorphisms were found between the parents for 70 of 117 SSR markers which amplified in the RILs. Single factor ANOVA analysis was used to identify QTL. Fatty acid modifier QTL were confirmed as follows: palmitic acid [LG D2 (Satt372), LG L (Satt166, Satt561 and Satt229)], stearic acid [LG B2 (Satt070, Satt556, Satt304 and Satt474), LG C2 (Satt557, Satt460, Satt079, Satt307), LG L (Satt481,

Satt1156, Satt076, and Satt166)], oleic acid [LG L (Satt076, Satt166, Satt527, Satt561, Satt229)], linoleic acid [LG F (Satt335), LG L (Satt076, Satt166, Satt527 and Satt561)], linolenic acid [LG L (Satt481, Satt156, Satt076, Satt166, Satt527 and Satt561)]

Introduction

Soybean [Glycine max (L.) Merrill] is an annual legume originally domesticated in China (ca.1700-1100 B.C.) and is now grown worldwide for its protein and oil (Hymowitz, 1987). Soybean oil primarily contains the following fatty acids: palmitic acid (16:0), stearic acid (18:0), oleic acid (18:1), linoleic acid (18:2), and linolenic acid (18:3) (Wilson, 2004). The concentration and characteristics of each fatty acid determine the quality of products produced. For example linolenic acid is prone to oxidation which turns the oil product rancid thus reducing its shelf life. Partial hydrogenation is used to reduce the levels of linolenic acid, thereby increasing the shelf life of the product (Lusas, 2004; Pantalone et al., 2004). Hydrogenation of soybean oil; however, results in the formation of trans-fatty acids, which are coronary disease risk factors. Consumer awareness and FDA rules (Federal Register, 1999) requiring food processors to report the amount of trans-fats in their products have provided an incentive for plant breeders to produce cultivars whose fatty acid profiles can be manipulated to produce the desired products. Low linolenic acid varieties, such as Vistive™ (Monsanto Company) which contains 3% linolenic acid, are available to processors. In addition, fatty acid profiles can be manipulated to increase the concentration of other fatty acids, such as stearic acid which is reported to have positive health benefits (Yu et al., 1995).

Hyten et al. (2004b) detected fatty acid modifier QTL in an Essex x 'Williams' RIL population. The purpose of this study is to confirm previously reported soybean fatty acid modifier QTL and submit confirmed QTL as cQTL to SoyBase (www.soybase.org).

Materials and Methods

Essex x Williams 82 Population Development

The initial crosses for the Essex x 'Williams 82' population were made at the University of Tennessee Plant Science Farm (KPSF) of the East Tennessee Research and Education Center (ETREC) in Knoxville, TN in the summer of 2002. The population was advanced from the F_2 to the F_6 generation through single seed descent (Brim, 1966)

Field Experiments

In 2006, 146 recombinant inbred lines (RILs) and the two parents, Essex and Williams 82 were planted in four reps of single plant hill plots (sown 3 beans/hill then thinned to one plant per plot about 14 days after emergence) in a randomized complete block Honeycomb design (Fasoulas and Fasoula, 1995). In addition, 129 genotypes were planted in 3.1m rows sown at 8 beans/0.31m for seed increase for the 2007 experiments. Lines which did not yield enough seed were further increased in winter nursery to ensure that there was sufficient seed for the 2007 experiments.

Based on maturity data collected in 2006, the population was divided by days to maturity into three population subsets: early (22 genotypes), mid (96 genotypes) and late (23 genotypes) separated from each other by ten days in maturity. Five genotypes were dropped from the study because of inadequate seed amounts. In 2007, each population subset which included the two parents was planted in two 6.1m row plots in a randomized complete block design replicated three times at the East Tennessee Research and Education Center (ETREC), the Highland Rim Research and Education Center (HRREC)

and the West Tennessee Research and Education Center (WTREC). Checks in the population subsets were assigned by maturity group as follows: Early (Macon and LD00-3309), mid (5002T) and late (5002T and 5601T). All checks used in this study corresponded with those used by the Northern and Southern uniform testing programs.

Phenotypic Traits

Five soybean pods were harvested from random plants within each plot. Pods were harvested mid-height from the main stem. This was done to prevent potential mechanical mixing of sample seed at harvest and to standardize the node height for fatty acid samples. Pods were threshed individually, and seed samples from each pod were compiled for each plot. Five seeds from a sample were used for fatty acid analysis. Seeds were crushed and 0.5mL/seed extraction solvent consisting of 8:5:2 (v:v:v) of chloroform, hexane, and methanol was added. Seeds were left to extract at room temperature for at least 6 hours. A 100 µL sample of the extracted oil supernatant was transferred into a 1.5 mL auto sampler vial to which 75 µL of a methylating reagent [0.5 mL sodium methoxide in methanol/petroleum ether, and ethyl ether(1:4:2 v:v:v)] and 0.75 mL hexane was added. The resulting fatty acid methyl ester was analyzed using a Hewlett Packard HP 6890 gas chromatograph (Palo Alto, CA) equipped with model 7673 autosampler, flame ionization detector, and an immobilized 30m long x 0.53 mm x 0.5µm inner diameter All-Tech AT-Silar capillary column with 0.5µL fused stationary phase. Operating conditions were: carrier He (20mL/Min), 20:1 (v:v) split injection, injection temperature 250ºC, detector temperature 275ºC and column temperature 240ºC.

The RM-1 standard, suitable for measuring soybean oil, was used to determine the relative fatty acid concentration of the test samples.

DNA Extraction and Polymerase Chain Reaction

In 2006, four plants each of 146 RILs and the two parents were grown in the growth chamber for leaf collection for DNA isolation. Leaves were harvested from all four plants at V4 stage. In addition, five leaves per genotype were randomly sampled from the first replication of plots in each test in ETREC in 2007 to ensure that we had enough leaves from which to extract DNA. All leaves were flash frozen in liquid nitrogen and then stored at -80 ºC until DNA was extracted.

DNA was extracted from the leaves of RILs and parents utilizing the Qiagen Plant DNeasy Extraction Kit (Qiagen, Valencia, CA). Polymerase chain reaction (PCR) consisted of 6.9 µL of ddH_2O, 4 µL Hotmaster TAQ (5Prime Gaithersburg, Md), 0.55 µL of 5 µM WellRED labeled forward primer (Proligo, Boulder, CO) and 0.55 µL of 5 µM reverse primer (Sigma-Genosys, Woodlands, TX), and 3 µL of 15 ng/µL template DNA. PCR reactions were performed in a 96-well MBS Hybaid thermocycler (Hybaid, Franklin, MA). PCR conditions consisted of initial denaturation at 95ºC for 2 min followed by 35 cycles of a) denaturation at 92ºC for 1 min, b) 47ºC for annealing for 1 min, c) 68ºC for 5 min for extension and, d) a final cycle at 72ºC for extension for 5 min. Parents were screened with a total of 203 $(ATT)_n$ type simple sequence repeat (SSR) genetic markers developed by Cregan et al. (1999) with the goal to use markers to confirm seed quality trait QTL. Markers were selected based on their proximity to QTL intervals previously reported by Hyten et al. (2004a, 2004b). One hundred and seventeen

markers were found to be polymorphic. Seventy of the 117 polymorphic SSR markers which amplified in the progeny were used to genotype RILs for QTL analysis. Sequence information for the SSRs is publicly available from Soybase (www.soybase.org; verified April 3, 2008.)

DNA Gel Electrophoresis

A Beckman-Coulter CEQ 8800 Genome Analysis System (Beckman-Coulter, Fullerton, CA) was used to separate PCR products by gel capillary electrophoresis. Beckman Coulter fragment analysis software was used to estimate PCR amplicon base pair lengths for detectable polymorphisms. Markers were considered to be polymorphic when the parental amplicon base pair lengths differed by ≥3 nucleotides. Polymorphic primers detected in the parents were used to genotype RILs. RILs were scored (1 = Essex allele, 2 = Heterozygote, and 3 = Williams 82 allele) for all markers including pubescence color and flower color. Heterozygote scores were dropped from the data set so as to only report potential additive genetic effects.

Data Analysis

The early, mid, and late population subsets were combined across the three environments (ETREC, HRREC and WTREC) to create three populations named set-1, set-2 and set-3 respectively. It was hypothesized that by separating the populations by maturity group we might be able to see QTL differences within a more narrow range of maturity. Sets 1, 2 and 3 were also combined to form the total population which could enable us to further detect QTL with greater detection power.

Phenotypic data was analyzed using the MIXED procedure of SAS (SAS Institute Inc., 2002) software to determine whether there were significant genotypic differences among lines within sets and within the total population. Environments and replication were considered as two blocking factors in the model.

QTL were analyzed via single factor ANOVA SAS GLM (SAS Institute Inc., 2002) procedure to determine if there were any associations between SSR markers and phenotypic least square means for the traits of interest. MAPMAKER /EXP 3.0 (Lander et al.1987; Lincoln et al.1992) was used to estimate distance between SSR markers. A minimum likelihood of odds (LOD) $\geq$3.0 and a maximum distance $\leq$ 50 cM were used to test linkage among markers. Marker order was similar to the soybean composite integrated map (Song et al., 2004) a finding that was also supported in the Essex x Williams population by Hyten et al. (2004a). The soybean composite integrated map was used to determine marker distances because it is based on five mapping populations and marker distances are more accurate, moreover, we were attempting to confirm QTL intervals reported by Hyten et al. (2004b) who used the composite integrated map in their calculations.

Heritability for each trait was calculated on an entry mean basis for three environments and three replications according to Nyquist (1991) whose formula was presented by Panthee et al. (2005).

REML estimation in PROC MIXED was used to determine the variance components for heritability calculations. PROC CORR (SAS 2002) was used to determine phenotypic correlations between traits.

Results and Discussion

Seventy SSR markers placed on linkage groups (LGs) B2, C2. D1a, D2, F, G, I, H, K, L and M were analyzed by single factor ANOVA for their possible association with least square means for palmitic acid, stearic acid, oleic acid, linoleic acid and linolenic acid content in the confirmation population.

There were significant ($p < 0.05$) genotypic differences for palmitic acid content among the recombinant inbred lines (RILs) in the confirmation population, in all population sets and in the total population. Palmitic acid content ranged from 115 to 150 g kg^{-1}, with a mean of 124 g kg^{-1} in the confirmation population (Table 4.1). Parental palmitic acid content was similar so the wide range of palmitic acid content observed in the RILs was due to transgressive segregation (based on ± 2 std. deviations from the high parent or low parent). The moderate heritability estimate (0.57) observed for palmitic acid in this population was due to additive genotypic effects because environmental and genotype by environment variation was not significant ($p \geq 0.05$), and because heterozygotes were excluded from the analysis.

Hyten et al. (2004b) detected modifier QTL for palmitic acid on LGs D2, K and L. The palmitic acid content QTL on LG D2 was confirmed by PAL5 (set-3) (Table 4.2). The palmitic acid content QTL previously detected by Hyten et al. (2004b) on LG K was not detected in this study. PAL16 (total population), PAL18 (set-3 and the total population), PAL19 (set-2 and the total population) and PAL20 (set-2) confirmed palmitic acid content QTL previously reported by Hyten et al. (2004b) on LG L.

Table 4.1 Descriptive statistics for fatty acid traits in an F_6 derived soybean RIL population from an Essex x Williams 82 cross grown in three environments in Tennessee

Trait	Min	Max	Mean	Std. Dev	Essex	Williams 82	h^2
Palmitic g Kg^{-1}	115	150	124	4.8	126	122	0.57
Stearic g Kg^{-1}	35	55	42	3.8	42	47	0.77
Oleic g Kg^{-1}	191	330	244	28.0	228	281	0.81
Linoleic g Kg^{-1}	440	570	531	24.4	543	498	0.79
Linolenic g Kg^{-1}	48	72	59	4.8	61	53	0.63

Table 4.2 QTL, map position and additive genetic effects for palmitic acid content in F_6 derived early, medium and late maturing soybean RIL subpopulations and total population from an Essex x Williams 82 cross grown in three environments in Tennessee

				Set-1			Set-2			Set-3			Total Population		
QTL[a]	**Marker[b]**	**LG[c]**	**Pos[d]**	**Effect[e]**	**P**	**R^2**	**Effect[e]**	**P**	**R^2**	**Effect[e]**	**P**	**R^2**	**Effect[e]**	**P**	**R^2**
PAL1	Satt070	B2	72.8	-	ns	-	3	0.0014	0.04	-	ns	-	2	0.0032	0.07
PAL2	Satt556	B2	73.21	-	ns	-	-	ns	-	-4	0.0036	0.14	-	ns	-
PAL3	Satt557	C2	112.19	-	ns	-	-	ns	-	-6	0.0199	0.14	-	ns	-
PAL4	Satt412	D1b	72.57	-	ns	-	-	ns	-	-5	0.0307	0.19	-	ns	-
PAL5	Satt372	D2	39.35	-	ns	-	-	ns	-	4	0.0057	0.15	-	ns	-
PAL6	Satt269	F	11.37	-4	0.0272	0.11	-	ns	-	-5	0.0004	0.22	-	ns	-
PAL7	Satt252	F	16.08	-	ns	-	-	ns	-	-4	0.0024	0.16	-2	0.0241	0.05
PAL8	Satt149	F	18.125	-3	0.0482	0.07	-2	0.0211	0.02	-	ns	-	-2	0.0079	0.06
PAL9	Satt114	F	63.69	-	ns	-	-	ns	-	3	0.023	0.09	-	ns	-
PAL10	Satt324	G	33.26	-5	0.0042	0.19	-	ns	-	-	ns	-	2	0.0259	0.04
PAL11	Satt317	H	89.52	-	ns	-	4	0.0048	0.04	-	ns	-	2	0.0299	0.05
PAL12	Satt292	I	82.78	-3	0.0418	0.07	-2	0.0185	0.02	-	ns	-	-2	0.0047	0.07
PAL13	Satt555	K	42.71	-	ns	-	-2	0.0436	0.02	-	ns	-	-	ns	-
PAL14	Satt481	L	54.57	4	0.0283	0.08	-	ns	-	-	ns	-	2	0.0466	0.03
PAL15	Satt076	L	61.35	-	ns	-	3	0.0003	0.05	-	ns	-	3	0.0001	0.13
PAL16	Satt166	L	66.55	4	0.0102	0.11	-	ns	-	4	0.0124	0.12	3	0.0005	0.1
PAL17	Satt527	L	70.36	3	0.0389	0.08	-	ns	-	-	ns	-	-	ns	-
PAL18	Satt561	L	71.44	-	ns	-	3	0.0257	0.02	3	0.0044	0.15	3	0.0042	0.08
PAL19	Satt229	L	93.89	-	ns	-	3	0.0001	0.06	-	ns	-	3	0.0001	0.13
PAL20	Satt240	M	35.85	-	ns	-	-3	0.0085	0.03	-	ns	-	-	ns	-

[a] PAL1 to PAL20 denote twenty palmitic acid content QTL detected

[b] SSR marker associated with the QTL by SF ANOVA

[c] Linkage group based on the integrated soybean genetic linkage map

[d] Marker position on linkage group based on the integrated soybean genetic linkage map

[e] Average change (g Kg^{-1}) with respect to Essex allele

PAL15 (set-2 and the total population) and PAL16 (set-1, set-2 and the total population) were also detected on LG L. PAL15 and PAL16 are approximately 5cM apart and 10cM upstream from the interval PAL17-PAL20, so they may represent a new QTL on LG L. There were QTL detection differences for palmitic acid content by maturity group in the confirmation population.

Hyten et al. (2004b) found that a major QTL (R^2 = 13.1 close to the stem determinancy locus Dt1 on LG L from Williams decreased palmitic acid content by 1.4 g kg^{-1}. The same was true in our population with the Wiliams 82 allele decreasing palmitic acid by as much as 4 g kg^{-1}. Both these findings disagree with Rebetzke et al. (1998) who found that the normal palmitic acid phenotype was more frequently associated with indeterminate growth habit. Hyten et al. (2004b) detected modifier QTL for palmitic acid on LGs D2, K and L. The palmitic acid content QTL on LG D2 was confirmed by PAL5 (set-3) (Table 4.2). The palmitic acid content QTL previously detected by Hyten et al. (2004b) on LG K was not detected in this study. PAL16 (total population), PAL18 (set-3 and the total population), PAL19 (set-2 and the total population) and PAL20 (set-2) confirmed palmitic acid content QTL previously reported by Hyten et al. (2004b) on LG L. PAL15 (set-2 and the total population) and PAL16 (Set-1, set-2 and the total population) were also detected on LG L. PAL15 and PAL16 are approximately 5cM apart and 10cM upstream from the interval PAL17-PAL20, so they may represent a new QTL on LG L.

Other palmitic acid QTL not previously reported by Hyten et al. (2004b) were detected on LGs B2, C2, D1b, F, G, H, I, L and M. PAL1 and PAL2 on LG B2 are less than 1cM apart, and likely constitute a single QTL. The PAL1-PAL2 interval is

approximately 20cM downstream from the closest palmitic acid content QTL on LG B2 reported in Soybase (Diers and Shoemaker 1992), so it probably represents a new QTL. PAL3 is a major palmitic acid QTL ($R^2 = 0.14$) on LG C2 is a new QTL detected in this study. PAL4 on LG D1b has previously been reported by Panthee et al. (2006). PAL6-PAL8 on LG F have not previously been reported. The interval PAL6-PAL8 is approximately 7 cM wide and likely represents a single QTL.

PAL9 is about 44 cM downstream from the PAL6-PAL8 interval and is likely another new QTL on LG F. PAL10 is 10cM downstream from a palmitic acid content QTL on LG G previously reported by Panthee et al. (2006) and might be a new QTL in this region. PAL11, PAL12 and PAL13 on LGs H, I and K respectively, have not previously been reported are probably new palmitic acid content QTL in this population. PAL20 on LG M is approximately 30cM upstream of the closest palmitic acid QTL reported (Li et al., 2002) and is probably a new QTL.

There were significant ($p < 0.05$) genotypic differences for stearic acid content among the recombinant inbred lines (RILs) in population sets 2 and 3 and in the total population. Genotypic variation among RILs in set-1 was not significant ($p < 0.05$) therefore single factor ANOVA to detect QTL was not performed. Stearic acid content ranged from 35 to 55 g kg^{-1}, with a mean of 42 g kg^{-1} in the confirmation population (Table 4.1). Parental stearic acid content was similar so the wide range of stearic acid content observed in the RILs was due to transgressive segregation (based on ± 2 std. deviations from the high parent or low parent). The moderate heritability estimate (0.77) observed for stearic acid in this population was due to additive genotypic effects because

environmental and genotype by environment variation were not significant ($p > 0.05$) and because heterozygotes were excluded from the analysis.

Hyten et al. (2004b) detected stearic acid modifier QTL on LGs B2, C2 and L. The SSR marker Satt070 on LG B2 was associated with stearic acid content in the original mapping study. In this study, STE2 (Satt070) was a confirmed QTL (Table 4.3). In addition, STE1 (set-2 and the total population), STE3 (set-2 and the total population) and STE4 (set-2, set-3 and the total population) were confirmed within 6cM of STE2.

STE5 (set-2 and the total population) is 12 cM downstream from STE4 and is likely a different QTL. STE2 was previously mapped 12.3 cM from the FAS locus (Spencer et al., 2002). The Fas locus is mapped at 85 cM on the integrated soybean map (Song et al., 2003). This would indicate that the QTL confirmed in this study lie very close to the *Fas* locus and that the interval STE4-STE5 contains the Fas locus. This suggests that mapping modifier loci has the potential to reveal novel mutant alleles; a concept that may be extended to discovery of genetic control for other traits.

STE6 and STE7 (total population) were confirmed QTL on LG C2. STE9 (total population) was detected but not confirmed because its p-value (0.0441) exceeds the confirmation threshold of $p < 0.01$ established by the Soybean Genetics Committee.

Other stearic acid content QTL detected in this study that were not reported by Hyten et al. (2004b) were on LGs D1a, D1b, F, G and K. STE10 (set-3) and STE11 (set-2) on LG D1b are new QTL detected in this study. The interval STE10-STE11 is less than 1cM wide and is most likely a single QTL. STE12, STE13 and STE14 confirmed in set-2 are new stearic acid QTL detected on LG D1b. STE12 is approximately 39 cM upstream from the interval STE13-STE14 and is probably a different QTL.

Table 4.3 QTL, map position and additive genetic effects for stearic acid content in F_6 derived early, medium and late maturing subpopulations and total population from an Essex x Williams 82 cross grown in three environments in Tennessee

					Set-1			Set-2			Set-3		Total Population		
QTL[a]	**Marker[b]**	**LG[c]**	**Pos[d]**	**Effect[e]**	**P**	**R^2**	**Effect[e]**	**P**	**R^2**	**Effect[e]**	**P**	**R^2**	**Effect[e]**	**QTL[a]**	**Marker[b]**
STE1	Satt304	B2	65.56	-	ns	-	2	0.0001	0.11	3	0.0493	0.06	3	0.0004	0.11
STE2	Satt070	B2	72.8	-	ns	-	2	0.0001	0.13	3	0.0215	0.08	3	0.0001	0.16
STE3	Satt556	B2	73.21	-	ns	-	3	0.0001	0.18	3	0.0144	0.09	4	0.0001	0.23
STE4	Satt474	B2	75.34	-	ns	-	2	0.0001	0.15	4	0.0009	0.17	3	0.0001	0.19
STE5	Satt534	B2	87..59	-	ns	-	2	0.0017	0.04	-	ns	-	3	0.0001	0.13
STE6	Satt557	C2	112.19	-	ns	-	-	ns	-	-	ns	-	-3	0.002	0.11
STE7	Satt460	C2	117.77	-	ns	-	-	ns	-	-	ns	-	-2	0.007	0.06
STE8	Satt079	C2	117.87	-	ns	-	-	ns	-	-	ns	-	-	ns	-
STE9	Satt307	C2	121.27	-	ns	-	-	ns	-	-	ns	-	-1	0.0441	0.03
STE10	Satt129	D1a	11.22	-	ns	-	-	ns	-	-2	0.0239	0.09	-	ns	-
STE11	Satt147	D1a	11.99	-	ns	-	-1	0.0322	0.02	-	ns	-	-	ns	-
STE12	Satt546	D1b	87.20	-	ns	-	-2	0.0084	0.03	-	ns	-	-	ns	-
STE13	Satt274	D1b	116.35	-	ns	-	1	0.0064	0.03	-	ns	-	-	ns	-
STE14	Satt459	D1b	118.61	-	ns	-	2	0.0001	0.09	-	ns	-	-	ns	-

Table 4.3 Continued.

					Set-1			Set-2			Set-3		Total Population		
QTL[a]	**Marker[b]**	**LG[c]**	**Pos[d]**	**Effect[e]**	**P**	**R^2**	**Effect[e]**	**P**	**R^2**	**Effect[e]**	**P**	**R^2**	**Effect[e]**	**QTL[a]**	**Marker[b]**
STE15	Satt149	F	18.13	-	ns	-	-	ns	-	3	0.0332	0.08	-	ns	-
STE16	Satt191	G	96.57	-	ns	-	-	ns	-	-	ns	-	-3	0.0206	0.09
STE17	Satt240	K	52.88	-	ns	-	1	0.0087	0.03	-	ns	-	-	ns	-
STE18	Satt273	K	56.62	-	ns	-	1	0.0213	0.02	-	ns	-	-	ns	-
STE19	Satt481	L	54.57	-	ns	-	-2	0.0027	0.04	-	ns	-	-2	0.01	0.06
STE20	Satt156	L	56.14	-	ns	-	-2	0.0009	0.05	-	ns	-	-2	0.0049	0.07
STE21	Satt076	L	61.35	-	ns	-	-2	0.0001	0.06	-	ns	-	-1	0.0149	0.05
STE22	Satt166	L	66.55	-	ns	-	-2	0.0003	0.06	-	ns	-	-3	0.0001	0.13
STE23	Satt527	L	70.36	-	ns	-	-2	0.0025	0.04	-	ns	-	-2	0.0027	0.08
STE24	Satt561	L	71.44	-	ns	-	-2	0.0001	0.09	-	ns	-	-3	0.0001	0.15

[a] STE1 to STE24 denote twenty four stearic acid content QTL detected
[b] SSR marker associated with the QTL by SF ANOVA
[c] Linkage group based on the integrated soybean genetic linkage map
[d] Marker position on linkage group based on the integrated soybean genetic linkage map
[e] Average change (g kg^{-1}) with respect to Essex allele

The interval STE13-STE14 is approximately 2cM wide and is probably a single QTL. The QTL STE15 (set-3), STE16 (total population) and STE17 (set-2) are new QTL on LGs F, G and K detected in this study.

There were significant ($p < 0.05$) genotypic differences for oleic acid content among RILs in the confirmation population, in all population sets and in the total population. There were significant ($p < 0.05$) genotypic differences for oleic acid content among RILs in the confirmation population, in all population sets and in the total population. Oleic acid content ranged from 191 to 330 g Kg^{-1}, with a mean of 244 g Kg^{-1} in the confirmation population (Table 4.1). The moderately high heritability estimate (0.81) observed for oleic acid in this population suggests that most of the phenotypic variation observed was due to genotypic effects. Hyten et al (2004b) detected modifier QTL for oleic acid content on LGs D1b and L. OLE12 (set-2) on LG D1b was detected but not confirmed in this study (Table 4.4). There were few markers placed on this linkage group in this study, which might explain our inability to confirm this QTL. OLE21-OLE25 confirmed oleic acid QTL detected on LG L. The large effects of QTL in this region may be due to the Dt1 locus (89.13 cM on the soybean integrated map). Furthermore, major maturity QTL ($R^2 \geq 0.10$) were detected in this study in the interval OLE19-OLE25. However, whereas maturity QTL in this interval were only detected in set-1, oleic acid content QTL were detected in all population sets and in the total population. This might suggest that maturity QTL on this LG are fixed in this population and that the Dt1 locus is segregating, hence the difference in the number of QTL detected.

Table 4.4 QTL, map position and additive genetic effects for oleic acid content in F_6 derived early, medium and late maturing soybean RIL subpopulations and total population from an Essex x Williams 82 cross grown in three environments in Tennessee

				Set-1			Set-2			Set-3			Total Population		
QTL[a]	**Marker[b]**	**LG[c]**	**Pos[d]**	**Effect[e]**	**P**	**R^2**	**Effect[e]**	**P**	**R^2**	**Effect[e]**	**P**	**R^2**	**Effect[e]**	**P**	**R^2**
OLE1	Satt304	B2	65.55	-	ns	-	-14	0.0042	0.04	-	ns	-	-	ns	-
OLE2	Satt070	B2	72.8	-	ns	-	-19	0.0001	0.08	-	ns	-	-15	0.008	0.06
OLE3	Satt556	B2	73.31	-	ns	-	-10	0.0387	0.02	-	ns	-	-	ns	-
OLE4	Satt474	B2	75.34	-	ns	-	-17	0.0003	0.06	-	ns	-	-	ns	-
OLE5	Satt557	C2	112.19	-66	0.0001	0.35	-	ns	-	-	ns	-	-48	0.0001	0.25
OLE6	Satt460	C2	117.77	-42	0.0018	0.16	-15	0.0234	0.02	-	ns	-	-23	0.0005	0.09
OLE7	Satt079	C2	117.87	-33	0.0175	0.09	-	ns	-	-	ns	-	-	ns	-
OLE8	Satt307	C2	121.27	-	ns	-	-	ns	-	-25	0.0036	0.12	-12	0.0187	0.04
OLE9	Satt202	C2	126.24	-	ns	-	-	ns	-	-26	0.0026	0.15	-12	0.0255	0.04
OLE10	Satt077	D1a	43.39	-32	0.0267	0.07	-10	0.0313	0.02	-	ns	-	-12	0.035	0.04
OLE11	Satt203	D1a	61.89	-54	0.0002	0.22	-	ns	-	-	ns	-	-	ns	-
OLE12	Satt412	D1b	72.57	-	ns	-	12	0.0371	0.03	-	ns	-	-	ns	-
OLE13	Satt372	D2	39.35	-	ns	-	-10	0.0156	0.02	-	ns	-	-	ns	-
OLE14	Satt649	F	5.361	-	ns	-	-10	0.0377	0.02	-	ns	-	-	ns	-

Table 4.4 Continued.

					Set-1			Set-2			Set-3		Total Population		
QTL[a]	**Marker[b]**	**LG[c]**	**Pos[d]**	**Effect[e]**	**P**	**R^2**	**Effect[e]**	**P**	**R^2**	**Effect[e]**	**P**	**R^2**	**Effect[e]**	**P**	**R^2**
OLE15	Satt335	F	77.70	-	ns	-	-	ns	-	26	0.0045	0.15	-	ns	-
OLE16	Satt324	G	33.26	-	ns	-	-	ns	-	-30	0.0016	0.19	-	ns	-
OLE17	Satt102	K	30.28	61	0.0004	0.31	-	ns	-	-	ns	-	-	ns	-
OLE18	Satt475	K	78.68	-	ns	-	-10	0.0149	0.02	-	ns	-	-	ns	-
OLE19	Satt481	L	54.57	-44	0.0005	0.2	-20	0.0001	0.09	-	ns	-	-25	0.0001	0.21
OLE20	Satt156	L	56.14	-38	0.032	0.11	-25	0.0001	0.12	-	ns	-	-28	0.0001	0.22
OLE21	Satt076	L	61.35	-59	0.0001	0.33	-32	0.0001	0.22	-	ns	-	-34	0.0001	0.4
OLE22	Satt166	L	66.55	-60	0.0001	0.37	-31	0.0001	0.2	-26	0.0237	0.1	-38	0.0001	0.47
OLE23	Satt527	L	70.36	-65	0.0001	0.37	-30	0.0001	0.18	-	ns	-	-34	0.0001	0.36
OLE24	Satt561	L	71.44	-60	0.0001	0.34	-34	0.0001	0.24	-28	0.0093	0.13	-40	0.0001	0.47
OLE25	Satt229	L	93.89	-	ns	-	-22	0.0001	0.09	-	ns	-	-21	0.0001	0.14
OLE26	Satt540	M	35.85	28	0.0339	0.07	-10	0.0154	0.02	-	ns	-	-	ns	-

[a] OLE1 to OLE26 denote twenty six oleic acid content QTL detected

[b] SSR marker associated with the QTL by SF ANOVA

[c] Linkage group based on the integrated soybean genetic linkage map

[d] Marker position on linkage group based on the integrated soybean genetic linkage map

[e] Average change (g Kg^{-1}) with respect to Essex allele

Other oleic acid QTL not previously reported by Hyten et al. (2004b) were detected on LGs B2, C2, D1a, D2, F, G and M. OLE1-OLE4 detected on LG B2 is within 5cM of a similar QTL reported by Diers and Shoemaker (1992). OLE5-OLE9 on C2 have not previously been associated with oleic acid. In this population, major maturity QTL were mapped to this region in population set-1 and in the total population. This interval is also very close to the E1 gene which strongly suggests that maturity has a large effect on oleic acid QTL mapped in this region of LG C2. OLE10 and OLE11, OLE13, OLE14 and OLE15, OLE16, OLE17, OLE18 and OLE26 on LGs D1a, D2, F, G, K and M respectively were new oleic acid QTL detected in this population.

There were significant ($p< 0.0001$) genotypic differences for linoleic acid content among RILs in all population sets and in the total population. Linoleic acid content ranged from 440 to 570 g Kg^{-1}, with a mean of 531 g Kg^{-1}among RILs (Table 4.1). Despite the large difference between parental means, there was transgressive segregation (based on ± 2 std. deviations from the high parent or low parent). The moderately high heritability estimates observed for linoleic acid suggest that much of the phenotypic variation observed was due to genotypic effects. However, genotype by environment (G × E) and environmental effects were significant ($p < 0.01$) in this population. Selection for linolenic acid would be difficult as one would have to take into account the significant E and G × E interaction.

Hyten et al. (2004b) detected linoleic acid content QTL on LGs F and L. LIN15 (set-3) confirmed the linoleic acid content QTL on LG F (Table 4.5). LIN15 (set-3) on LG F was detected but not confirmed ($0.01 < p < 0.05$). LIN21-LIN24 in sets-1, set-2 and the total population confirmed linoleic acid content QTL on LG L.

Table 4.5 QTL, map position and additive genetic effects for linolenic acid content in F_6 derived early, medium and late maturing soybean RIL subpopulations and total population from an Essex x Williams 82 cross grown in three environments in Tennessee

				Set-1			Set-2			Set-3			Total Population		
QTL[a]	**Marker[b]**	**LG[c]**	**Pos[d]**	**Effect[e]**	**P**	**R^2**	**Effect[e]**	**P**	**R^2**	**Effect[e]**	**P**	**R^2**	**Effect[e]**	**P**	**R^2**
LIN1	Satt304	B2	65.56	-	ns	-	8	0.0388	0.02	-	ns	-	-	ns	-
LIN2	Satt070	B2	72.80	-	ns	-	11	0.0028	0.04	-	ns	-	-	ns	-
LIN3	Satt474	B2	75.34	-	ns	-	11	0.0064	0.03	-	ns	-	-	ns	-
LIN4	Satt557	C2	112.19	66	0.0002	0.32	-	ns	-	-	ns	-	44	0.0001	0.28
LIN6	Satt079	C2	117.77	36	0.0099	0.1	-	ns	-	-	ns	-	-	ns	-
LIN5	Satt460	C2	117.77	44	0.0016	0.17	11	0.0254	0.02	-	ns	-	20	0.0004	0.09
LIN7	Satt307	C2	121.27	-	ns	-	-	ns	-	22	0.0027	0.13	11	0.0118	0.05
LIN8	Satt202	C2	126.24	-	ns	-	-	ns	-	22	0.004	0.13	10	0.0265	0.04
LIN9	Satt077	D1a	43.39	-	ns	-	9	0.0312	0.02	-	ns	-	10	0.0381	0.04
LIN10	Satt203	D1a	61.89	51	0.0008	0.19	-	ns	-	-	ns	-	-	ns	-
LIN11	Satt412	D1b	72.57	-	ns	-	-10	0.0341	0.03	-	ns	-	-	ns	-
LIN12	Satt546	D1b	87.2	-	ns	-	8	0.0414	0.02	-	ns	-	-	ns	-
LIN13	Satt649	F	5.36	-	ns	-	9	0.0199	0.02	-	ns	-	-	ns	-
LIN14	Satt114	F	63.69	-	ns	-	-	ns	-	-19	0.0272	0.09	-	ns	-
LIN15	Satt335	F	77.7	-	ns	-	-	ns	-	-26	0.0018	0.17	-	ns	-

Table 4.5 Continued.

					Set-1			Set-2			Set-3		Total Population		
QTL[a]	**Marker[b]**	**LG[c]**	**Pos[d]**	**Effect[e]**	**P**	**R^2**	**Effect[e]**	**P**	**R^2**	**Effect[e]**	**P**	**R^2**	**Effect[e]**	**P**	**R^2**
LIN16	Satt324	G	33.26	-	ns	-	-	ns	-	26	0.0028	0.17	-	ns	-
LIN17	Satt292	I	82.78	-	ns	-	9	0.0093	0.03	-	ns	-	-	ns	-
LIN18	Satt475	K	78.68	-	ns	-	8	0.0302	0.02	-	ns	-	-	ns	-
LIN19	Satt481	L	54.57	43	0.0011	0.18	16	0.0001	0.09	-	ns	-	21	0.0001	0.19
LIN20	Satt156	L	56.14	38	0.0422	0.09	22	0.0001	0.13	-	ns	-	23	0.0001	0.2
LIN21	Satt076	L	61.35	59	0.0001	0.31	26	0.0001	0.2	-	ns	-	28	0.0001	0.35
LIN22	Satt166	L	66.55	56	0.0001	0.32	27	0.0001	0.23	-	ns	-	33	0.0001	0.46
LIN23	Satt527	L	70.36	61	0.0001	0.35	26	0.0001	0.19	-	ns	-	29	0.0001	0.33
LIN24	Satt561	L	71.44	58	0.0001	0.3	28	0.0001	0.25	23	0.019	0.11	33	0.0001	0.43
LIN25	Satt229	L	93.89	-	ns	-	15	0.0001	0.07	-	ns	-	16	0.0005	0.1
LIN26	Satt540	M	35.85	-	ns	-	8	0.0346	0.02	-	ns	-	-	ns	-

[a] LIN1 to LIN26 denote twenty six linoleic acid content QTL detected

[b] SSR marker associated with the QTL by SF ANOVA

[c] Linkage group based on the integrated soybean genetic linkage map

[d] Marker position on linkage group based on the integrated soybean genetic linkage map

[e] Average change (g Kg^{-1}) with respect to Essex allele

Major maturity QTL were mapped in this region in this study, which might explain the large effects and R^2 values observed. LIN25 (set-2 and the total population) detected in this study is approximately 4 cM downstream from the Dt1 locus.

Other linoleic acid QTL on LGs B2, C2, D1a, D1b, G, I, K, and M were new QTL detected in this population.

There were significant ($p < 0.0001$) genotypic differences for linolenic acid content among RILs in all population sets and in the total population. Linolenic acid content for the RILs ranged from 48 to 72 g Kg^{-1}, with a mean of 59 g Kg^{-1}(Table 4.1). Parental means for linolenic acid content were similar, so the large difference observed among RILs was due to transgressive segregation (based on ± 2 std. deviations from the high parent or low parent).

There were significant ($p < 0.0001$) genotypic differences for linolenic acid content among RILs in all population sets and in the total population. Heritability for linolenic acid was moderate, and the significant ($p < 0.0001$) G ×E and E interaction observed in this population would make it difficult to select for linoleic acid.

Hyten et al. (2004b) detected QTL for linolenic acid on LGs F and L. The QTL on LG F in the interval 7.9-17.4 cM was not detected in this population even though there were sufficient markers placed in this interval at approximately one marker every 4cM (Table 4.6). LEN17 (set-1, set-2 and the total population), LEN18 (total population), LEN19-LEN20 (set-1, set-2 and the total population) confirmed the linolenic acid content QTL in the interval 42.6-62.2 cM on LG L.

Table 4.6 QTL, map position and additive genetic effects of for linolenic acid content in F_6 derived early, medium and late maturing soybean RIL subpopulations and total popualtion from an Essex x Williams 82 cross grown in three environments in Tennessee

				Set-1			Set-2			Set-3			Total Population		
QTL[a]	**Marker[b]**	**LG[c]**	**Pos[d]**	**Effect[e]**	**P**	**R^2**	**Effect[e]**	**P**	**R^2**	**Effect[e]**	**P**	**R^2**	**Effect[e]**	**P**	**R^2**
LEN1	Satt070	B2	72.8	-	ns	-	3	0.0185	0.02	-	ns	-	2	0.014	0.05
LEN2	Satt474	B2	75.34	-	ns	-	3	0.0426	0.02	-	ns	-	-	ns	-
LEN3	Satt534	B2	87.59	-	ns	-	-	ns	-	-	ns	-	-2	0.0398	0.04
LEN4	Satt557	C2	112.19	7	0.0212	0.14	-	ns	-	7	0.0309	0.12	6	0.0006	0.14
LEN5	Satt460	C2	117.77	-	ns	-	-	ns	-	-	ns	-	4	0.005	0.06
LEN6	Satt307	C2	121.27	-	ns	-	-	ns	-	3	0.0033	0.13	3	0.0073	0.06
LEN7	Satt202	C2	126.24	-	ns	-	-	ns	-	4	0.0037	0.14	2	0.0347	0.03
LEN8	Satt203	D1a	61.89	6	0.0386	0.08	-	ns	-	-	ns	-	-	ns	-
LEN9	Satt546	D1b	87.20	-	ns	-	3	0.0473	0.02	-	ns	-	2	0.0234	0.05
LEN10	Satt372	D2	39.35	-	ns	-	3	0.0104	0.03	-	ns	-	3	0.0159	0.05
LEN11	Satt335	F	77.70	-	ns	-	-	ns	-	-5	0.0008	0.19	-	ns	-
LEN12	Satt144	F	102.08	-	ns	-	-	ns	-	-4	0.0231	0.09	-	ns	-

1Table 4.6 Continued

					Set-1			Set-2			Set-3		Total Population		
QTL[a]	**Marker[b]**	**LG[c]**	**Pos[d]**	**Effect[e]**	**P**	**R^2**	**Effect[e]**	**P**	**R^2**	**Effect[e]**	**P**	**R^2**	**Effect[e]**	**P**	**R^2**
LEN13	Satt324	G	33.26	-	ns	-	-	ns	-	4	0.0057	0.15	-	ns	-
LEN14	Satt191	G	96.57	-	ns	-	4	0.045	0.05	-	ns	-	3	0.0225	0.09
LEN15	Satt102	K	30.28	-8	0.0131	0.17	-	ns	-	-	ns	-	-	ns	-
LEN16	Satt475	K	78.68	-	ns	-	3	0.0168	0.02	-	ns	-	2	0.0148	0.05
LEN17	Satt481	L	54.57	8	0.0026	0.15	4	0.001	0.05	-	ns	-	4	0.0001	0.19
LEN18	Satt156	L	56.14	7	0.0335	0.11	3	0.0127	0.03	-	ns	-	4	0.0001	0.17
LEN19	Satt076	L	61.35	8	0.0012	0.17	4	0.0001	0.08	-	ns	-	5	0.0001	0.29
LEN20	Satt166	L	66.55	8	0.0002	0.22	4	0.0022	0.04	3	0.0409	0.08	5	0.0001	0.29
LEN21	Satt527	L	70.36	8	0.0006	0.2	4	0.0001	0.07	-	ns	-	6	0.0001	0.31
LEN22	Satt561	L	71.44	7	0.001	0.17	4	0.0001	0.07	3	0.0416	0.08	5	0.0001	0.32
LEN23	Satt229	L	98.89	3	0.0023	0.04	-	ns	-	-	ns	-	4	0.0001	0.13

[a] LEN1 to LEN23 denote twenty three linolenic acid content QTL detected
[b] SSR associated with the QTL by SF ANOVA
[c] Linkage group based on the integrated soybean genetic linkage map
[d] Marker position on linkage group based on the integrated soybean genetic linkage map
[e] Average change (g Kg^{-1}) with respect to Essex allele

LEN21-LEN22 (set-1, set-2 and the total population) and LEN23 (set-1 and the total population) confirmed the linolenic acid QTL in the interval 72.5-91.1 cM on LG L. QTL detected on linkage group L were also associated with maturity in this population.

Other QTL for linolenic acid QTL detected in this population were not reported by Hyten et al. (2004b). The QTL on LGs B2, C2, D1a, D1B, D2, F, G and K were novel in this population. LEN16 on LG K has previously been reported (Diers and Shoemaker, 1992).

Conclusion

Most of the fatty acid modifier QTL mapped by Hyten et al. (2004b) were confirmed in this study [e.g. palmitic acid (PAL5, PAL20), stearic acid (STE2, STE6), oleic acid (OLE21-OLE25), linoleic acid (LIN21-Lin24) and linolenic acid (LEN17 and LEN18)]. QTL conditioning for stearic acid on LG were confirmed very close to the Fas locus. These QTL could be useful in selecting for stearic acid.

QTL confirmed on LG C2 (STE6, STE7) and L (PAL19, PAL20, OLE21-OLE25, LIN21-LIN24, and LEN17-LEN23 were also associated with major ($R^2 > 0.10$) maturity QTL. Maturity QTL in this population were mostly detected in population set-1 which might suggest that this population is no longer segregating for maturity. Even, though we were able to observe a difference in trait expression because population subsets were divided by days to maturity, the small population size could also influence the patterns seen. To adequately document the pleiotropic effects of maturity on fatty acid modifiers a larger population that is segregating for maturity would be much more informative.

In addition, QTL detected on LG L (PAL19, OLE25, LIN25), may have been influenced by the Dt1 locus. Fatty acid modifier QTL which also seem to be pleiotropic for maturity and growth habit may prove difficult to use in marker assisted selection, as crosses would call for large parental differences in maturity and growth habit.

References

References

Brim, C. A., 1966. A modified pedigree method of selection in soybeans. Crop Sci. 6, 220.

Brummer, E. C., G. L. Graef, J. Orf, J. R. Wilcox, and R. C. Shoemaker, 1997. Mapping QTL for seed protein and oil content in eight soybean populations. Crop Sci. 37, 3 70-378.

Cregan, P. B., T. Jarvik, A. L. Bush, R. C. Shoemaker, K. G. Lark, A. L. Kahler, N. Kaya, T. T. VanToai, D. G. Lohnes, and J. Chung, 1999. An integrated genetic linkage map of the soybean genome. Crop Sci. 39, 1464—1490.

Csanadi, G., J. Vollmann, G. Stift, and T. Lelley. 2001. Seed Quality QTLs identified in a molecular map of early maturing soybean. Theor. Appl. Genet. 103: 912-919.

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

Fasoula, Vasilia A., Donna K. Harris and H. Roger Boerma. 2004. Validation and designation of quantitative trait loci for seed protein, seed oil, and seed weight from two soybean populations. Crop Sci. 44: 1218-1225.

Fasoulas, A.C and V.A Fasoula 2005. Honey comb selection designs. *In* Jules Janick (ed.) Plant Breeding Reviews. Vol. 13. John Wiley and Sons Inc. New York, USA.

Fehr, W.R., Caviness, C.E., Burmond, D.L., and Pennington, J.S. 1971. Stages of development decriptions for soybeans for soybeans, Glycine max (L.) Merrill. Crop Sci. 11. 929-931.

Hyten, D. L. QTL mapping and identification of GxE interactions of agronomic and seed quality Traits in soybean. 2002.

Hyten, D. L., V. R. Pantalone, C. E. Sams, A. M. Saxton, D. Landau-Ellis, T. R. Stefaniak, and M. E. Schmidt, 2004a. Seed quality QTL in a prominent soybean population. Theor. Appl. Genet. 109: 552—561.

Hyten, D. L., Vincent R. Pantalone, Arnold M. Saxton, Michael E. Schmidt and Carl E. Sams. 2004b. Molecular mapping and identification of soybean fatty acid modifier quantitative trait loci. J. Am. Oil Chem. Soc. 81 (12) :1115-1118.

Lee, S. H., M. A. Bailey, M. A. R. Mian, E. R. Shipe, D. A. Ashley, W. A. Parrott, R. S. Hussey, and H. R. Boerma, 1996. Identification of quantitative trait loci for plant height, lodging, and maturity in a soybean population segregating for growth habit. Theor. Appl. Genet. 92, 516-523.

Lincoln, S. E., M. J. Daly, and E. S. Lander, 1993. Constructing Genetic Linkage Maps with MAPMAKER, Version 3. A Tutorial and Reference Manual. Release 3rd ed. Whitehead Institute for Biomedical Research, Cambridge, MA.

Nyquist, W. E., 1991. Estimation of heritability and prediction of selection response in plant populations. Crit. Rev. Plant Sci. 10, 235-322.

Orf, J. H., K. Chase, T. Jarvik, L. M. Mansur, P. B. Cregan, F. R. Adler, and K. G. Lark, 1999b. Genetics of soybean agronomic traits. I. Comparison of three related recombinant inbred populations. Crop Sci. 39, 1642-1651.

Pantalone, Vincent R., David R. Walker, Ralph Dewey and Istvan Rajcan. 2004. DNA Marker-Assisted Selection for Improvement of Soybean Oil Concentration and Quality. p. 284-285, 301-303. *In* H. Thomas Stalker and E. Charles Brummer (eds.) Legume and crop Genomics. AOCS Press, Champaign, Illinois.

Panthee, D. R., V. R. Pantalone, D. R. West, A. M. Saxton, and C. E. Sams.2005. Quantitative Trait Loci for Seed Protein and Oil Concentration, and Seed Size in Soybean. Crop Sci. 45: 2015-2022.

Panthee D.R, V. R. Pantalone, A. M. Saxton. 2006. Modifier QTL for fatty acid composition in soybean oil. Euphytica 152: 67-73.

SAS Institute Inc. 2002. SAS Guide for Personal Computers, 9th edn. SAS Institute, Cary, NC.

Song, Q. J., L. F. Marek, R. C. Shoemaker, K. G. Lark, V. C. Concibido, X. Delannay, J. E. Specht, and P. B. Cregan, 2004. A new integrated genetic linkage map of the soybean. Theor. Appl. Genet. 109, 122-128.

Specht, J.E., D.J. Hume, and S.V. Kumudini. 1999. Soybean yield potential- a genetic and physiological perspective. Crop Sci. 39 :1560–1570.

Specht, J. E., K. Chase, M. Macrander, G. L. Graef, J. Chung, J. P. Markwell, M. Germann, J. H. Orf, and K. G. Lark, 2001. Soybean response to water. A QTL analysis of drought tolerance. Crop Sci. 41, 493-509.

SAS Institute. 2002. The SAS system for Windows. Release 9.0. SAS Inst., Cary, NC.

Soystats. [online] Available at www.soystats.com (verified 14 April 2008).

Wilcox J.R. 1998. Increasing seed protein in soybean with eight cycles of recurrent selection. Crop Sci.38: 1536-1540.

Wilcox, James R. 2004. World Distribution and Trade of Soybean. *In* H.R Boerma and J. E. Specht (eds.) Soybeans: Improvement, Production, and Uses, 3rd ed. Agron. Monogr. 16. ASA, CSSA, SSSA, Madison, WI.

Wilcox, J. R. Sixty Years of Improvement in Publicly Developed Elite Soybean Lines. 2001. Crop Sci. 41: 1711-1716.

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

Vita

Catherine Nyaguthii Nyinyi was born in Nyeri, Kenya on the 9th of March 1982. She attended Murray State University in Murray, Kentucky and received a Bachelor of Science degree in Agriculture in 2004. In 2005, she worked as a corn breeding intern for Monsanto Company in Arlington, WI.

In 2006 she enrolled at the University of Tennessee and began soybean breeding research under Dr. Vincent Pantalone. Catherine hopes to continue her education at the University of Tennessee pursuing a PhD. in plant science with research in soybean breeding.