5-2012

Detection of Quantitative Trait Loci for Marker-Assisted Selection of Soybean Isoflavone Genistein

Christopher Joseph Smallwood
csmallw1@utk.edu

Recommended Citation
Smallwood, Christopher Joseph, "Detection of Quantitative Trait Loci for Marker-Assisted Selection of Soybean Isoflavone Genistein."
https://trace.tennessee.edu/utk_gradthes/1208

This Thesis is brought to you for free and open access by the Graduate School at Trace: Tennessee Research and Creative Exchange. It has been accepted for inclusion in Masters Theses by an authorized administrator of Trace: Tennessee Research and Creative Exchange. For more information, please contact trace@utk.edu.
To the Graduate Council:

I am submitting herewith a thesis written by Christopher Joseph Smallwood entitled "Detection of Quantitative Trait Loci for Marker-Assisted Selection of Soybean Isoflavone Genistein." 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:

Dean A. Kopsell, Carl E. Sams, Dennis R. West

Accepted for the Council:

Dixie L. Thompson

Vice Provost and Dean of the Graduate School

(Original signatures are on file with official student records.)
Detection of Quantitative Trait Loci
for Marker-Assisted Selection
of Soybean Isoflavone Genistein

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

Christopher Joseph Smallwood
May 2012
Acknowledgements

I would like to extend my gratitude to all those who helped me throughout the course of this Master’s Thesis. Foremost, I would like to thank Dr. Vincent Pantalone, who has served as my major advisor. Thank you for the excellent educational opportunities you have provided me with during this program. Also, thank you for the guidance and support you have provided me with in conducting this research.

I would also like to thank my committee members, Dr. Dean Kopsell, Dr. Carl Sams, and Dr. Dennis West. Thank you for your support and counsel in the completion of this thesis.

Much assistance was provided by many former and current members of the Soybean Breeding Team in the completion of this research. To all of you who helped, I thank you: Dr. Catherine Nyinyi, Greg Allen, Ben Fallen, Deborah Landau-Ellis, Elizabeth Meyer, Nicole Tacey, Ben Wiggins, and Suzannah Wiggins.

For assistance with data analysis I would like to thank Dr. Arnold Saxton, and for assistance with genotypic data I would like to thank Dr. Perry Cregan and Dr. David Hyten.

I would like to extend special thanks to Dr. Fred Allen for originally suggesting that I attend graduate school, as well as for support and guidance during the course of this research.

I am very grateful for support to pursue this research from the Tennessee Soybean Promotion Board and the University of Tennessee Institute of Agriculture.

Finally, I would like to thank my family for their support and encouragement. Especially, I would like to thank my wife Jennifer, who has supported my efforts and inspired me to a higher level of performance.
Abstract

Soybean [Glycine max (L.) Merrill] is an important crop throughout the world. Among the many seed quality traits contained in soybean are isoflavones, which are associated with numerous health benefits, including cancer prevention, improved cardiovascular health, improved bone health, and reduced menopausal symptoms. This study sought to identify quantitative trait loci (QTL) controlling soybean isoflavones genistein, daidzein, glycine, and total isoflavone content to gain a better understanding of genetic regions controlling production of these compounds. The phenotypic data for QTL detection was generated in 2009 from a population of 274 recombinant inbred lines (RILs) separated into three field tests based on maturity (early, mid, and late) and grown in three locations (Knoxville, TN; Harrisburg, IL; and Stuttgart, AR). Genotypic data was obtained using 1,536 single nucleotide polymorphism (SNP) markers, of which 480 were polymorphic. Overall, 21 QTL were detected for soybean isoflavones, including 7 for genistein, 5 for daidzein, 3 for glycine, and 6 for total isoflavones. Of these 21 QTL, 8 were newly detected, while 13 were validated from previous studies. Marker-assisted selections (MAS) were made using the QTL for genistein, which is typically the most abundant isoflavone, for comparison with phenotypic selections. Challenges exist when considering MAS for quantitative traits such as isoflavones, including concerns with epistatic interactions and genotype × environment interactions. However, isoflavone improvement with MAS would be useful as phenotyping data is costly and time consuming. Comparisons of MAS and phenotypic selection methods were done in 2010 and 2011 in field tests grown in three locations (Knoxville, TN; Springfield, TN; Milan, TN). Results from this study indicate that phenotypic selections outperformed MAS for genistein. However MAS for genistein did show improvements in relation to parental lines, as well as unselected RILs included in field tests for comparison. Additionally, genistein was significantly correlated with other isoflavones, as well as with yield. More research should be done as the costly and time consuming process of collecting phenotypic data for isoflavones provides incentive to pursue MAS as an improvement strategy.
# Table of Contents

CHAPTER 1

Introduction and Literature Review
- Introduction ................................................................. 1
- Literature Review ......................................................... 5
- Objectives ...................................................................... 10
- References .................................................................. 11
  References .................................................................. 12

CHAPTER 2

Validation and Detection of quantitative trait loci for Soybean Isoflavones
- Abstract ........................................................................ 19
- Introduction .................................................................. 19
- Material and Methods .................................................. 21
  Plant Materials .............................................................. 21
  Agronomic Traits .......................................................... 22
  Gas Chromatography .................................................... 23
  Near Infrared Reflectance Spectroscopy ......................... 23
  DNA Extraction and Molecular Analysis ......................... 24
  Data Analysis ............................................................... 24
- Results and Discussion ................................................ 25
  Phenotypic Traits ........................................................ 25
  Genetic Mapping and QTL Detection ............................ 27
- Conclusion .................................................................... 31
- References .................................................................. 33
  References .................................................................. 34

CHAPTER 3

Comparison of MAS and Phenotypic Selection Methods for Soybean Isoflavone Genistein
- Abstract ........................................................................ 39
- Introduction .................................................................. 39
- Material and Methods .................................................. 41
  Plant Materials .............................................................. 41
  Agronomic Traits .......................................................... 43
  Gas Chromatography .................................................... 43
  Near Infrared Reflectance Spectroscopy ......................... 44
  DNA Extraction and Molecular Analysis ......................... 45
  QTL Detection and Marker-Assisted Selections .................. 45
  Data Analysis ............................................................... 46
- Results and Discussion ................................................ 47
- Conclusion .................................................................... 50
- References .................................................................. 52
  References .................................................................. 53

Chapter 4

Conclusion and Future Research
- Conclusion and Future Research .................................... 57
Appendix A: tables ........................................................................................................... 60
Appendix B: Figures ........................................................................................................... 74
Vita ..................................................................................................................................... 93
List of Tables

Table 2.1 Descriptive statistics and heritability values for a population of 274 RILs derived from ‘Essex’ and ‘Williams 82’, grown over three environments (Knoxville, TN; Harrisburg, IL; and Stuttgart, AR) with three replications. Parental LSMEANS also included. ........................................................................................................ 61

Table 2.2 Phenotypic correlations between genistein, daidzein, glycitein, and total isoflavones with agronomic and seed quality traits of interest in soybean from a population of 274 RILs derived from ‘Essex’ and ‘Williams 82’, grown over three environments (Knoxville, TN; Harrisburg, IL; and Stuttgart, AR) with three replications. ............................................................................................................ 62

Table 2.3 Quantitative trait loci associated with the isoflavone genistein in individual maturity tests in an ‘Essex’ × ‘Williams 82’ Population of 274 RILs grown over three environments (Knoxville, TN; Harrisburg, IL; and Stuttgart, AR). GEN1-GEN7 represent 7 QTL for the isoflavone genistein. .......................................................... 63

Table 2.4 Quantitative trait loci associated with the isoflavone daidzein in individual maturity tests in an ‘Essex’ × ‘Williams 82’ Population of 274 RILs grown over three environments (Knoxville, TN; Harrisburg, IL; and Stuttgart, AR). DAI1-DAI5 represent 5 QTL for the isoflavone daidzein. ............................................................. 64

Table 2.5 Quantitative trait loci associated with the isoflavone glycitein in individual maturity tests in an ‘Essex’ × ‘Williams 82’ Population of 274 RILs grown over three environments (Knoxville, TN; Harrisburg, IL; and Stuttgart, AR). GLY1-GLY3 represent 3 QTL for the isoflavone glycitein. .................................................. 65

Table 2.6 Quantitative trait loci associated with total isoflavones in individual maturity tests in an ‘Essex’ × ‘Williams 82’ Population of 274 RILs grown over three environments (Knoxville, TN; Harrisburg, IL; and Stuttgart, AR). ISO1-ISO6 represent 6 QTL for total isoflavones. .......................................................... 66

Table 3.1 Quantitative trait loci associated with height in individual maturity tests in an ‘Essex’ × ‘Williams 82’ Population of 274 RILs grown over three environments (Knoxville, TN; Harrisburg, IL; and Stuttgart, AR). HGT1 represents 1 QTL for soybean height. ............................................................. 67
Table 3.2 Comparison of height LSMEANS for lines in the 4eMAS, 4LMAS, and 5eMAS field tests from 2010 and 2011. ................................................................. 68

Table 3.3 Comparison of height LSMEANS for tail selections in the 4LMAS and 5eMAS field tests from 2010 and 2011. .......................................................................................... 69

Table 3.4 Comparison of genistein LSMEANS for lines in the 3L, 4e, and 5 Late maturity groups from 2010 and 2011. ........................................................................................................... 70

Table 3.5 Comparison of genistein LSMEANS for tail selections in the 3L, 4e, 4L, and 5e Maturity Groups from 2010 and 2011................................................................. 71

Table 3.6 Comparison of genistein for Lines in the 5e Maturity Group from 2010 and 2011................................................................................................................................. 72

Table 3.7 Phenotypic correlations between genistein and agronomic and seed quality traits of interest in soybean from LSMEANS of 41 RILs derived from ‘Essex’ and ‘Williams 82’ grown in seven field tests (Gen3L, Gen4e, Gen4L, Gen5e, Gen4eMAS, Gen4LMAS, and Gen5eMAS). Field tests were grown in three environments (Knoxville, TN; Springfield, TN; and Milan, TN) with three replications over two years (2010 and 2011). ........................................................................ 73
List of Figures

Figure 1.1 Chemical structures of soybean isoflavone aglycones (Pan et al., 2001)........ 75
Figure 1.2 Structural similarity between isoflavones and estrogen (Setchell and Cassidy, 1999). ........................................................................................................................................ 76
Figure 2.1 Distributions of genistein and daidzein means from 274 F$_5$ derived RILs of ‘Essex’ × ‘Williams 82’ separated into individual tests based on maturity, with each test containing 91 or 92 RILs. Field tests were grown in 2009 in Knoxville, TN; Harrisburg, IL; and Stuttgart, AR. All plots were harvested at maturity............... 77
Figure 2.2 Distributions of glycitein and total isoflavone means from 274 F$_5$ derived RILs of ‘Essex’ × ‘Williams 82’ separated into individual tests based on maturity, with each test containing 91 or 92 RILs. Field tests were grown in 2009 in Knoxville, TN; Harrisburg, IL; and Stuttgart, AR. All plots were harvested at maturity. ........ 78
Figure 2.3 Genetic map of ‘Essex’ × ‘Williams 82’ population of 274 RILs mapped with 480 polymorphic SNP markers. Chromosome 13 is Split Into 13a and 13b. ............ 79
Figure 2.4 QTL positions for genistein (shown in red) from Early, Mid, and Late tests of 274 RILs of ‘Essex’ × ‘Williams 82’ grown over three environments (Knoxville, TN; Harrisburg, IL; and Stuttgart, AR). E1 maturity locus (chromosome 6, shown in green) and Dt1 growth habit locus (chromosome 19, shown in blue). The cM positions are from this population. The consensus map 4.0 positions of each BARC SNP are available at http://www.soybase.org/................................................................. 80
Figure 2.5 QTL positions for daidzein (shown in red) from Early, Mid, and Late tests of 274 RILs of ‘Essex’ × ‘Williams 82’ grown over three environments (Knoxville, TN; Harrisburg, IL; and Stuttgart, AR). E1 maturity locus (chromosome 6, shown in green) and Dt1 growth habit locus (chromosome 19, shown in blue). The cM positions are from this population. The consensus map 4.0 positions of each BARC SNP are available at http://www.soybase.org/................................................................. 81
Figure 2.6 QTL positions for glycitein (shown in red) from Early, Mid, and Late tests of 274 RILs of ‘Essex’ × ‘Williams 82’ grown over three environments (Knoxville, TN; Harrisburg, IL; and Stuttgart, AR). E1 maturity locus (chromosome 6, shown in
green). The cM positions are from this population. The consensus map 4.0 positions of each BARC SNP are available at http://www.soybase.org/................. 82

Figure 2.7 QTL positions for total isoflavones (shown in red) from Early, Mid, and Late tests of 274 RILs of ‘Essex’ × ‘Williams 82’ grown over three environments (Knoxville, TN; Harrisburg, IL; and Stuttgart, AR). E1 maturity locus (chromosome 6, shown in green) and Dt1 growth habit locus (chromosome 19, shown in blue). The cM positions are from this population. The consensus map 4.0 positions of each BARC SNP are available at http://www.soybase.org/............................................. 83

Figure 2.8 Diagram of the phenylpropanoid pathway (Gutierrez-Gonzalez et al., 2010a). Enzymes used in pathway: phenylalanine ammonia lyase (PAL), cinnamic acid 4-hydroxylase (C4H), 4-coumarate:CoA ligase (4CL), chalcone synthase (CHS), chalcone reductase (CHR), chalcone isomerase (CHI), isoflavone synthase (IFS), flavanone 3-hydroxylase (F3H), dihydroflavonol reductase (DFR) and isoflavone reductase (IFR)......................................................................................................... 84

Figure 3.1 Selected markers for height MAS for candidate RILs from the 2009 early test (92 F5 derived RILs) grown over three environments (Knoxville, TN; Harrisburg, IL; and Stuttgart, AR). LOD-1 interval used for selection (shown in red) and Dt1 growth habit locus (shown in blue) also displayed. The cM positions are from this population. The consensus map 4.0 positions of each BARC SNP are available at http://www.soybase.org/.......................................................... 85

Figure 3.2 Selected markers for height MAS for candidate RILs from the 2009 mid test (91 F5 derived RILs) grown over three environments (Knoxville, TN; Harrisburg, IL; and Stuttgart, AR). LOD-1 interval used for selection (shown in red) and Dt1 growth habit locus (shown in blue) also displayed. The cM positions are from this population. The consensus map 4.0 positions of each BARC SNP are available at http://www.soybase.org/.......................................................... 86

Figure 3.3 Selected markers for height MAS for candidate RILs from the 2009 late test (91 F5 derived RILs) grown over three environments (Knoxville, TN; Harrisburg, IL; and Stuttgart, AR). LOD-1 interval used for selection (shown in red) and Dt1 growth habit locus (shown in blue) also displayed. The cM positions are from this
population. The consensus map 4.0 positions of each BARC SNP are available at http://www.soybase.org/. ................................................................. 87

Figure 3.4 Selected markers for genistein MAS for candidate RILs from the 2009 early test (92 F5 derived RILs) grown over three environments (Knoxville, TN; Harrisburg, IL; and Stuttgart, AR). LOD-1 interval used for selection (shown in red) and E1 maturity locus (shown in green) also displayed. A QTL was also detected on chromosome 13b in the early test, but it was not used for selections as no markers were within the LOD-1 interval. The cM positions are from this population. The consensus map 4.0 positions of each BARC SNP are available at http://www.soybase.org/. ................................................................. 88

Figure 3.5 Selected markers for genistein MAS for candidate RILs from the 2009 mid test (91 F5 derived RILs) grown over three environments (Knoxville, TN; Harrisburg, IL; and Stuttgart, AR). LOD-1 interval used for selection (shown in red) and Dt1 growth habit locus (shown in blue) also displayed. The cM positions are from this population. The consensus map 4.0 positions of each BARC SNP are available at http://www.soybase.org/. ................................................................. 89

Figure 3.6 Selected markers for genistein MAS for candidate RILs from the 2009 late test (91 F5 derived RILs) grown over three environments (Knoxville, TN; Harrisburg, IL; and Stuttgart, AR). LOD-1 interval used for selection (shown in red), E1 maturity locus (shown in green), and Dt1 growth habit locus (shown in blue) also displayed. The cM positions are from this population. The consensus map 4.0 positions of each BARC SNP are available at http://www.soybase.org/. ………………… 90

Figure 3.7 Comparison of LSMEAN heights of all tail selections in the 4LMAS and 5eMAS Field Tests from 2010 and 2011. Tukey-Kramer mean separation; LSMEANS followed by the same letter within a test are not significantly different at the 5% level of probability. ……………………………………………………………… 91

Figure 3.8 Comparison of genistein for LSMEAN tail selections in the 3L, 4e, and 4L Maturity Groups from 2010 and 2011. Tukey-Kramer mean separation; LSMEANS followed by the same letter within a test are not significantly different at the 5%
level of probability. Tail selections from MAS tests shown in red and phenotypic
tests shown in blue.
CHAPTER 1
INTRODUCTION AND LITERATURE REVIEW
Introduction

Soybean \([\text{Glycine max (L.) Merrill}]\) is an economically important agricultural crop in many countries throughout the world. Initially domesticated in China (ca. 1500-1100 B.C.), soybeans are now grown worldwide for many different purposes. In its early history, soybean remained primarily an Asian crop. After being introduced to Europe in the 1500-1700s A.D., soybean was first brought to the U.S. in 1765. Initially grown in the eastern U.S., soybean was moved to Illinois in the 1850s and from there quickly spread to the rest of the Corn Belt (Hymowitz, 2004). The range for soybean has not changed much in the U.S. since the 1850s, with most of the hectarage planted east of the Rocky Mountains where there is adequate precipitation to support widespread production (Wilcox, 2004).

Soybean production has increased significantly in recent decades. During the period from 1985 to 2002, soybean production in the U.S. increased by 49% (Wilcox, 2004). In 2010, over 31 million hectares of soybeans were planted in the U.S., yielding over 90 million metric tons (www.soystats.com verified February 28, 2012). The estimated value for soybeans planted in the U.S. in 2010 exceeded 37 billion dollars (www.nass.usda.gov verified February 28, 2012; www.soystats.com verified February 28, 2012). Soybean production has also increased significantly in recent decades in other countries with suitable climate and available land. In 2010, the U.S. was the leading producer of soybeans, followed by Brazil, Argentina, China, and India (www.soystats.com verified February 28, 2012).

This increase in soybean production is largely due to the high demand, favorable production costs, and quality seed traits. Traditionally, oil and protein have been the primary seed quality traits of interest for soybean. Currently, soybean is produced in greater abundance than any other oilseed throughout the world (Wilcox, 2004; www.soystats.com verified February 28, 2012). Soybean oil consists of five primary fatty acids, which occur in a relative abundance of 10% for palmitic (16:0), 4% for stearic (18:0), 22% for oleic (18:1), 54% for linoleic (18:2), and 10% for linolenic (18:3) (Wilson, 2004). Because soybean oil is used significantly in many human food products, the fatty acid profile is of high importance. Most of the fatty acid composition consists of unsaturated fats, of which oleic acid is the most desirable due to its increased stability and favorable cardiovascular effects. Reduction in polyunsaturated fatty acids is desirable, in particular linolenic acid (Panthee et al., 2006), which requires hydrogenation to improve stability. Furthermore, reduction in palmitic acid is desirable as it is a saturated fatty
acid with unfavorable cardiovascular effects. With the goal of improving the fatty acid profile of soybean oil for human consumption, breeding efforts have been initiated to increase the percentage of oleic acid and decrease the percentages of palmitic and linolenic acids (Fehr, 2007).

Another use for soybean oil is the production of biodiesel fuel. Concerns over high oil prices, limited global oil supply, and dependence on the Middle East for oil production have inspired efforts to find renewable sources of energy. This has prompted the Federal government to pass the Energy Independence and Security Act of 2007, as well as subsequent legislation requiring substantial increases in biofuel production. Biodiesel produced from soybean oil has emerged as a potential renewable energy source to help alleviate these concerns. From 1999 to 2008, biodiesel consumption in the U.S. rose from 500,000 gallons to 690 million gallons, decreasing back to 315 million gallons in 2010 (www.soystats.com verified February 28, 2012). The fatty acid profile is an important consideration for biodiesel produced from soybean oil. Soybeans with a higher concentration of oleic acid can improve the oxidative stability of biodiesel (Fallen et al., 2011). Because increased oleic acid is also desirable for human consumption, efforts are already underway to increase oleic acid content in soybean oil (Fehr, 2007). As a N-fixing legume, soybean requires little N fertilization. Compared to other crops used in biodiesel production, soybean requires relatively low energy input to grow. For this reason, soybean is the only crop used in biodiesel production that has a low net energy yield and a high-energy ratio (de Vries et al., 2010).

Soybean seed consists of approximately 40% protein on a dry weight basis. The protein portion of the soybean seed is used considerably in livestock feed. For non-ruminant livestock, soybean meal is the most commonly used protein supplement due to its dependable supply and well balanced amino acid profile (Min et al., 2009). Increased protein content is an important goal for soybean breeders (Durham, 2003). However, soybean meal as a protein supplement can also be problematic for non-ruminant livestock such as poultry and swine. Soybeans contain a high concentration of phytic acid, which is difficult to digest for poultry and swine. The phytic acid is passed along in the manure of these animals, which creates a P pollution problem (Raboy, 2002). In an effort to alleviate this concern, breeders have begun efforts to reduce the phytic acid contained in soybeans (Walker et al., 2006; Scaboo et al., 2009).
Soybean is used significantly as a human food source as well. Potential health benefits associated with soybean consumption have sparked interest in other quality traits from soybean, in particular, isoflavones. Soybean isoflavones are phytochemicals synthesized via the phenylpropanoid pathway (Bennett et al., 2004) with several benefits to the plant; including stimulation of soil microbe rhizobium in root nodule formation for N-fixation, and anti-fungal and anti-pathogenic activity (Ogbuewu, et al., 2010). The isoflavones found in soybean are genistein, daidzein, and glycine (Figure 1.1). Each of the isoflavones can exist as aglycones, glucosides, or glucoside derivatives, including; 6”-O-malonyl-esters and 6”-O-acetyl-esters (Wilson, 2004). Soybean isoflavones occur most abundantly in the malonylglucoside forms: 6”-O-malonylgenistein, 6”-O-malonyldaidzin, and 6”-O-malonylglycitin (Charron et al., 2005). Upon ingestion, however, the malonylglucosides, along with the acetylglucosides, are converted to the glucoside forms; which are subsequently metabolized into the aglycone forms (Brouns, 2002). Isoflavones are structurally similar to the hormone estrogen, and can act as estrogen mimics in humans (Figure 1.2) (Setchell, 1998).

In Asia, diets high in soybean isoflavones have been common for centuries. Soybean has been grown for thousands of years in Asia as a primary source of nutrition, anchored by such food products as tofu, miso, tempeh, and soy sauce (Hymowitz, 2004). Western diets, in contrast, typically contain much lower levels of soy isoflavones (Brouns, 2002). Recent research has indicated that typical western diets may be more likely to influence the development of certain types of cancer, as well as coronary heart disease, when compared with vegetarian or semi-vegetarian diets in some Asian countries (Adlercreutz, 1995). Many of the health benefits associated with Asian diets are attributed to soybean isoflavones. While more research is needed, many studies show soybean isoflavones may play an important role in cancer prevention (Birt et al., 2001). Consumption of soybean isoflavones can reduce the risk of coronary heart disease by improving blood pressure and total cholesterol in high-risk middle aged men (Sagara et al., 2003). Problems with diabetes and obesity may also be reduced by soybean isoflavones (Bhathena and Velasquez, 2002). Additional health benefits associated with soybean isoflavones include anti-oxidant activity, improved bone health, improved cardiovascular health, and decreased menopausal symptoms (Brouns, 2002).

As isoflavone content has become more important to consumers, soybean breeders have worked to improve seed levels. In addition to the associated health benefits, soybean isoflavones
are positively correlated to several other important soybean traits. A study by Morrison et al. (2008) showed that breeders working to improve soybean yield also improved isoflavone levels, albeit unknowingly. This study showed a positive correlation between yield and total isoflavone content (Morrison et al., 2008). In observing this, it should be noted that irrigation, which has a strong association with yield improvement, can increase isoflavone levels in soybean (Bennett et al., 2004). Primomo et al. (2005) showed a significant positive correlation between total isoflavone content and yield, maturity, plant height, and lodging, while showing a negative correlation between total isoflavone content and protein. Negative correlations between isoflavones and linolenic acid, as well as isoflavones and protein content have been noted by Wilson (2004). Additionally, Meng et al. (2011) observed an association between increased isoflavone content and soybean resistance to aphids.

Predicting isoflavone content in soybeans can be difficult because of environmental variability (Eldridge and Kwolek, 1983; Gutierrez-Gonzalez et al., 2009). Some of the environmental factors influencing isoflavone content include temperature and moisture (Lozovaya et al., 2005) as well as soil (Barion et al., 2010). This variability can make it difficult to select for isoflavone improvement solely based on phenotypic data. In spite of these challenges, improving soybean isoflavone content is an important objective for soybean breeders.

**Literature Review**

Numerous studies have shown soybean isoflavones to be a quantitatively inherited trait (Meksem et al., 2001, Kassem et al., 2004, Kassem et al, 2006, Primomo et al., 2005, Zeng et al., 2009). In order to gain a better understanding of the genetic regions that control isoflavone expression, identification and mapping of quantitative trait loci (QTL) for isoflavones is an important goal of soybean breeders. Researchers working to create genetic maps for soybean have used several different DNA markers in order to detect allelic polymorphisms. Some of the different markers used include older techniques, such as restriction fragment length polymorphisms (RFLPs), and also more recent methods, such as simple sequence repeats (SSRs or microsatellites) and single nucleotide polymorphisms (SNPs) (Torres et al., 2010). The genetic map created by Cregan et al. (1999) using 606 SSR markers resulted in a consensus of 20 linkage groups (LGs) on the soybean genome, corresponding to the haploid chromosome number. Updated genetic maps have been created for the soybean genome in recent years (Song
et al., 2004). The most recent map was created using 1,536 SNP markers, and determined that the soybean genome is approximately 2,300 centimorgans (cM) in length (Hyten et al., 2010). The creation of genetic linkage maps for soybeans have proved useful for identifying QTL (Hyten et al., 2010).

Once identified, QTL can be used to make marker assisted selections (MAS) for genetic improvement. Important considerations when using MAS include predicting QTL which are stable in multiple environments and using a large population size to predict QTL (Bernardo, 2008). Assuming these considerations are met, there are several advantages in using MAS as a selection method, including increased reliability, improved use of time, and reduced cost in comparison with conventional screening methods (Torres et al., 2010). The improvement in reliability is of particular interest when selecting for isoflavones, which have been shown to vary greatly by environment (Gutierrez-Gonzalez et al., 2010; Murphy et al., 2009; Eldridge and Kwolek, 1983).

Differences in environment are not the only source of variation for soybean isoflavones. Research has shown that genetic effects, along with the interaction between environmental and genetic effects, are also significant causes of isoflavone variation (Gutierrez-Gonzalez et al., 2009). In order to gain a better understanding of the genetic component of isoflavone expression, researchers have begun to identify and map QTL controlling these traits. A study conducted by Meksem et al. (2001) developed a population of 100 recombinant inbred lines (RILs) from using ‘Essex’ and ‘Forrest’ as the parent lines. With interval mapping (IM), 107 polymorphic SSR markers were used to create a genetic linkage map for this population. Seven QTL were identified on 5 different LGs. Three of the identified QTL were for glycitein, and they were on LGs B1 (chromosome 11), H (chromosome 12), and N (chromosome 3). Three of the QTL were for daidzein, and they were on LGs N (chromosome 3), K (chromosome 9), and A1 (chromosome 5). One of the QTL was for both glycitein and genistein, located on LG K (chromosome 9). Broad sense heritability estimates for daidzein (79%), genistein (22%), and glycitein (88%) were calculated for this population. The high heritability estimates for daidzein and glycitein suggest that much of the variation within this population was due to genetic effects.

Following up on the previous study, Kassem et al. (2004) used 240 polymorphic SSR markers to genotype the same population of 100 RILs derived from the ‘Essex’ × ‘Forrest’ cross using IM. Six of the QTL from the previous study were confirmed. The additional SSR markers
allowed for the detection of 2 new QTL; 1 for genistein and daidzein, located on LG B2 (chromosome 14) and 1 for glycitein, located on LG D1a+Q (chromosome 1). Kassem et al. (2006) revisited the ‘Essex’ × ‘Forrest’ population using the more accurate composite interval mapping (CIM) technique. Only 2 of the previously reported 8 QTL for isoflavones were confirmed, while 14 new QTL were reported. The resulting 16 QTL were located on 6 different LGs: B1 (chromosome 11), A2 (chromosome 8), D1a (chromosome 1), N (chromosome 3), M (chromosome 7), and G (chromosome 18) (Kassem et al., 2006).

A similar study was conducted by Primomo et al. (2005) using both IM and CIM. A population of 207 RILs were developed from a cross between AC756 (low isoflavone parent) and RCAT (high isoflavone parent). The RILs were grown in two different environments in Ontario, Canada. A genetic linkage map was created using 99 polymorphic SSR markers. Seventeen QTL were identified for genistein, daidzein, glycitein, and total isoflavones on 9 different LGs: A1 (chromosome 5), C2 (chromosome 6), D1a (chromosome 1), F (chromosome 13), G (chromosome 18), H (chromosome 12), J (chromosome 16), K (chromosome 9), and M (chromosome 7). Five of the QTL identified by Primomo et al. (2005) were located in similar genomic regions to QTL previously identified in the ‘Essex’ × ‘Forrest’ population (Kassem et al., 2004). These 5 QTL, located on LGs A1 (chromosome 5), D1a (chromosome 1), H (chromosome 12), K (chromosome 9), and N (chromosome 3), are of considerable interest because they were identified in highly different environments using different genetic material. Broad sense heritability for genistein, daidzein, glycitein, and total isoflavones ranged from 35% to 50%, which was quite different from the values obtained in the ‘Essex’ × ‘Forrest’ population (Meksem et al., 2001). Of additional interest was the identification of 23 epistatic interactions for isoflavone content by Primomo et al. (2005). The identification of epistatic interactions indicates that QTL effects may include genetic components beyond additive and dominant gene action. The possible influence of epistatic interactions should be considered when using MAS for isoflavones (Primomo et al., 2005).

The importance of detecting QTL using diverse germplasm in varying environments is evident due to the genotypic and environmental effects, as well as the effects of the genotype x environment interaction on isoflavone content (Primomo et al., 2005). With this in mind, Zeng et al. (2009) sought to discover QTLs controlling isoflavone content in a population of 136 RILs derived from a cross between two Chinese developed soybean cultivars; ‘Zhongdou 27’ (high
isoflavone parent) and ‘Jiunong 20’ (low isoflavone parent). This population of RILs was grown in seven different environments in China. A genetic map was created using 99 polymorphic SSR markers. The mapping technique used was the same used by Primomo et al. (2005). Eleven QTL on 9 different LGs: F (chromosome 13), I (chromosome 20), K (chromosome 9), A2 (chromosome 8), C2 (chromosome 6), M (chromosome 7), O (chromosome 10), D2 (chromosome 17), G (chromosome 18); were identified for genistein, daidzein, glycine, and total isoflavone content. Of particular interest was the QTL for genistein, glycine, and total isoflavones located on LG M (chromosome 7) at Satt 540. This QTL, which was determined to be significant in multiple locations in that study, had previously been found to be significant across multiple locations by Primomo et al. (2005). No epistatic interactions were found by Zeng et al. (2009). Broad sense heritability estimates ranged from 31% to 57%, which were similar to the values obtained by Primomo et al. (2005).

In an effort to uncover epistatic interactions for soybean isoflavones, Gutierrez-Gonzalez et al. (2009) conducted a study analyzing 196 RILs derived from a cross between ‘Essex’ and PI 437654. The RILs were grown in two different locations in Missouri. A genetic map was created using 276 polymorphic SSR and amplified fragment length polymorphism (AFLP). Several mapping techniques were used, including IM, CIM, mixed composite interval mapping (MCIM), and mixed interval mapping (MIM). Twenty-six QTL displaying additive epistatic interactions were uncovered for soybean isoflavones. These interactions varied greatly by environment, indicating that the degree of epistatic interaction is heavily influenced by the growing environment. Broad sense heritability was estimated to be 83%, 89%, 43%, and 86% for genistein, daidzein, glycine, and total isoflavone content, respectively. Most of the isoflavones displayed high heritability estimates, which was similar to results found by Meksem et al. (2001), indicating that several soybean isoflavones are primarily genetically controlled (Gutierrez-Gonzalez et al. 2009).

Further studies have continued to confirm previously identified QTL and discover new QTL for soybean isoflavones (Gutierrez-Gonzalez et al., 2010; Yoshikawa et al., 2010; Liang et al., 2010; Gutierrez-Gonzalez et al., 2011; Yang et al., 2011). Due to the high degree of isoflavone variation caused by genetic effects (Gutierrez-Gonzalez et al. 2009), QTL that have been identified could be useful for MAS in soybean (Primomo et al., 2005; Zeng et al., 2009). However, there are currently no studies that test the effectiveness of MAS for soybean
isoflavones. One possible reason for this could be the perceived difficulty of using MAS for a trait that is controlled by many genes with a large complexity of epistatic interactions (Gutierrez-Gonzalez et al. 2010).

It should be noted that plant breeders have successfully used MAS for other soybean traits (Concibido et al., 1996; Walker et al., 2002; Walker et al., 2004; Maroof et al., 2008; Sebastian et al., 2010), as well as for other economically important crops (Perumalsamy et al., 2010). Concibido et al. (1996) demonstrated the effectiveness of using MAS for soybean cyst nematode (SCN) resistance. This study compared the accuracy of selecting for SCN resistance based on conventional scoring techniques with the accuracy of using molecular markers to select for SCN resistance. The results showed a similar degree of accuracy using either technique. However, the cost and time for using MAS was determined to be lower than for using conventional screening techniques (Concibido et al., 1996).

Walker et al. (2002 and 2004) used MAS to successfully incorporate insect resistance into a population of BC2F3 plants (2002) and BC2F3 lines (2004) derived from a cross between ‘Jack-Bt’ (recurrent parent) and PI229358 (donor parent). These results indicate that MAS can be an effective tool in selecting for insect resistance. Additionally, MAS may be the only method of effectively pyramiding genes whose effects may be masked by the presence of other genes (Walker et al., 2002). Maroof et al. (2008) demonstrated another example of the value of MAS in conferring pest resistance. In that study, MAS was successfully used to pyramid three resistance genes to soybean mosaic virus (Maroof et al., 2008).

Sebastian et al. (2010) tested the possibility of using molecular markers to select for yield. Five separate elite mother line populations were used for this study. The mother line populations were considered to be heterogeneous because each was derived from a single seed in either the F3 or the F4 generation. QTLs for yield were identified within each of the populations and used to make selections. In field trials repeated over multiple years and locations, the yield of the bulked selections from each population was compared to yield of their respective mother population. Three of the selected lines yielded significantly higher than their respective mother line. All 5 of the selected lines were higher yielding than their respective mother lines, but two of them were not significant. The results from that study are especially encouraging, because, like isoflavones, yield is a highly complex trait influenced by both genotype and environment.
Objectives

The objectives of this study are to: (i) detect and validate QTL for soybean isoflavones in a random population of RILs derived from a cross between ‘Essex’ × ‘Williams 82’; (ii) compare the MAS method to the phenotypic selection method for the isoflavone genistein; and (iii) determine phenotypic correlations between genistein and other soybean isoflavones, as well as other traits of interest, to determine what effect genistein selections will have on other traits.
References
References


Soystats. [online] Available at www.soystats.com (verified February 28, 2012)


Transgressive segregation of isoflavone contents under the control of four QTLs in a cross between distantly related soybean varieties. Breeding Sci. 60:243-254.

Zeng, G. L., D. M. Li, Y. P. Han, W. L. Teng, J. Wang, L. Q. Qiu, and W. B. Li. 2009.

Abstract

Interest in soybean \textit{[Glycine max (L.) Merrill]} isoflavones has increased in recent years due to numerous potential health benefits. Analytical measurement methods for soybean isoflavones can be time consuming and costly. Consequently, QTL detection for marker assisted breeding is being examined for its potential for genetic gains. This study sought to detect QTL for soybean isoflavones in three different maturity tests (early, mid, and late) in a population of 274 recombinant inbred lines (RILs) derived from parental lines ‘Essex’ and ‘Williams 82’. The field tests were grown in a randomized complete block design (RCBD) replicated three times in three environments in 2009 (Knoxville, TN; Harrisburg, IL; and Stuttgart, AR). The population was genotyped with 1,536 single nucleotide polymorphism (SNP) markers, of which 480 were polymorphic. Phenotypic data for each replicate was collected with near infrared reflectance (NIR) spectroscopy. The equation used for NIR prediction of genistein, daidzein, and glycine was derived from high performance liquid chromatography (HPLC) analysis using 497, 499, and 492 samples with $R^2$ values of 0.85, 0.85, and 0.65, respectively. NIR is much faster and less expensive than the more commonly used HPLC method for isoflavone measurement. Each maturity test, containing 91 or 92 RILs, was analyzed separately for QTL. In total 21 QTL were detected: 7 for genistein (chromosomes 5, 6, 9, 9, 13, 17, and 19), 5 for daidzein (chromosomes 5, 6, 9, 13, and 19), 3 for glycine (chromosomes 6, 9, and 20), and 6 for total isoflavones (chromosomes 5, 5, 6, 9, 13, and 19). Of these 21 QTL, 8 QTL had not been previously reported, while 13 QTL were validated from other studies. The utilization of these QTL could potentially lead to marker-assisted selection approaches for genetic gains for soybean isoflavones.

Introduction

Soybean isoflavones have gained considerable interest in recent years as a potential benefit to human health. The isoflavones found in soybean are genistein, daidzein, and glycine. Soybean isoflavones are structurally similar to the hormone estrogen (Setchell, 1998), and can exist as aglycones, glucosides, or glucoside derivatives, including; 6”-0-malonyl-esters and 6”-0-acetyl-esters (Wilson, 2004). After synthesis via the phenylpropanoid pathway (Bennett et al., 2004), soybean isoflavones exist predominantly in the malonylglucoside form (Charron et al., 2005), and, upon ingestion, are converted into aglycones (Brouns, 2002).
Numerous benefits associated with soybean isoflavones include cancer prevention (Birt et al., 2001), reduced risk for coronary heart disease (Sagara et al., 2003), reduced problems with diabetes and obesity (Bhathena and Velasquez, 2002), improved bone and cardiovascular health, and decreased menopausal symptoms (Brouns, 2002). The soybean plant also benefits in several areas from isoflavones, including stimulation of soil microbe rhizobia in root nodule formation for N-fixation, and anti-fungal and anti-pathogenic activity (Ogbuewu, et al., 2010). Soybean isoflavones have been positively correlated with other important soybean traits. Among the traits positively associated with isoflavones are yield (Primomo et al., 2005; Morrison et al., 2008) and aphid resistance (Meng et al., 2011), while linolenic acid, an undesirable fatty acid found in soybean oil, has been negatively associated with isoflavone content (Wilson, 2004). Increasing soybean isoflavone content may become an important goal for plant breeders because of the large number of associated benefits.

Improving soybean isoflavone content can be difficult due to their quantitative inheritance (Meksem et al., 2001, Kassem et al., 2004, Kassem et al, 2006, Primomo et al., 2005, Zeng et al., 2009). Heritability estimates for soybean isoflavones are inconsistent, ranging from moderate (Primomo et al., 2005; Zeng et al., 2009) to high (Gutierrez-Gonzalez et al. 2009). Gaining better understanding of the genetic regions that control isoflavone content would be beneficial for making improvements. Toward this end, several studies have sought to identify quantitative trait loci (QTL) which control soybean isoflavone content. Meksem et al. (2001) identified 7 QTL for genistein, daidzein, and glycitein on 5 chromosomes in a population of 100 recombinant inbred lines (RILs) derived from the parental lines ‘Essex’ and ‘Forrest’. Following up on that study, Kassem et al. (2004) used an increased number of molecular markers on the same population of 100 RILs, and confirmed 6 of the previous QTL, while detecting 2 new QTL. Kassem et al. (2006) revisited the ‘Essex’ × ‘Forrest’ population using the composite interval mapping (CIM) technique and only confirmed 2 QTL from the previous study, while detecting 14 new QTL. Primomo et al. (2005) detected 17 QTL for genistein, daidzein, glycitein, and total isoflavones on 9 chromosomes. Of particular interest were 5 QTL located on similar genetic regions to those already detected by Kassem et al. (2004), who used a different population grown in different environments. In a similar study, Zeng et al. (2009) detected 11 QTL for genistein, daidzein, glycitein, and total isoflavones on 9 chromosomes. That population of 136 RILs derived from the Chinese cultivars ‘Zhongdou 27’ (high isoflavone parent) and ‘Jiunong 20’
(low isoflavone parent) was grown in 7 different environments in China. Of particular interest was a QTL detected for genistein, glycitein, and total isoflavones on chromosome 7, which was significant across multiple locations in China, and significant across multiple locations in Canada (Primomo et al., 2005). Additional QTL have been detected by Gutierrez-Gonzalez et al. (2009; 2010b; 2011), who have examined major and minor QTL for soybean isoflavones, as well as epistatic interactions.

Further challenges for soybean isoflavone improvement include variation by environment (Gutierez-Gonzalez et al., 2009; Murphy et al., 2009) and year (Eldridge and Kwolek, 1983). Environmental factors influencing isoflavone content include temperature and moisture (Lozovaya et al., 2005) as well as soil (Barion et al., 2010). These challenges reinforce the need to select QTL which are stable across multiple target environments (Bernardo, 2008). Finding a cheaper and faster measurement technique than high performance liquid chromatography (HPLC) for soybean isoflavones is also important in order to improve the efficiency of isoflavone selections among the large populations of individuals typically encountered in plant breeding programs. Another important consideration for soybean isoflavones is QTL detection based on maturity, as genistein, daidzein, and total isoflavones have been positively correlated with maturity (Primomo et al., 2005). Use of a large RIL population and large number of molecular markers are also important for QTL detection, in order to increase the likelihood of finding markers closely linked to QTL. Once detected, QTL can be used for marker-assisted selections (MAS). There are several advantages in using MAS as a selection method, including increased reliability, improved use of time, and reduced cost in comparison with conventional screening methods for some traits where that technology is effective (Torres et al., 2010).

Considering these factors, the objectives of the current research were to: (i) identify isoflavone QTL stable over multiple environments based on maturity; (ii) validate or confirm isoflavone QTL from previous studies using the near infrared reflectance (NIR) measurement technique; and (iii) obtain phenotypic correlations between isoflavones and other important agronomic and seed quality traits for soybean.

**Material and Methods**

**Plant Materials**

A population of 274 RILs was developed from the parental lines ‘Essex’ and ‘Williams 82’. The ‘Essex’ and ‘Williams 82’ cultivars used in this population were obtained from the
USDA soybean germplasm collection (www.ars-grin.gov), and a random single plant was intentionally selfed for two generations to provide highly homozygous parental lines to be crossed for RIL development. The initial cross was made in the summer of 2005 at the East Tennessee Research and Education Center (ETREC) in Knoxville, TN. The F₁ single plants were harvested in the fall of 2005 and grown as F₁ single plants in Puerto Rico at the Tropical Agricultural Research Station (TARS) in Isabela, Puerto Rico, in the winter of 2005-06. Following the single seed descent method (Brim, 1966), this population was advanced from the F₂ to the F₅ generation as follows: the F₂ seed harvested from TARS were grown at ETREC in the summer of 2006 and F₂:3 seed were harvested in the fall of 2006; the F₂:3 seed were grown at ETREC in the summer of 2007 and F₃:₄ seed were harvested in the fall of 2007; the F₃:₄ seed were grown at TARS in the winter of 2007-2008, F₄:₅ seed were harvested at maturity and grown at TARS in the spring of 2008; the F₅:₆ plants harvested from TARS were grown in Knoxville in the summer of 2008 as RILs in 3.1 m rows for agronomic data collection and leaf collection for DNA extraction. The F₅:₇ RILs harvested in the fall of 2008 were sent to Homestead, FL, for seed increase in the winter of 2008-2009. Seed harvested from the spring of 2009 were grown as 274 F₅:₈ RILs in the summer of 2009 tested in replicated yield trials. The 274 RILs were divided into three different tests based on maturity from the 2008 data (early, mid, and late), with each test containing 91 or 92 RILs. Both parents, along with four other check cultivars or lines were included in each test. The check lines for the early test (‘IA4004’, LD00-2817P, LD00-3309, and ‘Macon’), mid test (TN05-4008, TN06-189, TN06-196, and ‘5002T’), and late test (JTN-5203, ‘Osage’, ‘5002T’, and ‘5601T’), were selected to correspond with the appropriate maturity group for their respective test. The 2009 field trials were grown using a randomized complete block design (RCBD) replicated three times and grown in three different locations (Knoxville, TN; Harrisburg, IL; and Stuttgart, AR). Each entry was planted in a plot consisting of two adjacent rows 6.1 m in length, with the rows spaced 0.8 m apart.

**Agronomic Traits**

For the 2009 growing season, flower color was recorded at the R2 growth stage when 95% of the plants in a plot were in full bloom. Pubescence color and date of maturity were recorded during the R8 growth stage when 95% of the pods in a plot showed their mature color (Fehr and Caviness, 1977). Also at the R8 growth stage, height measurements were taken and lodging estimates were recorded. Height for each plot was estimated to be the average height of
the plants within the plot. Lodging estimates were taken on a 1 (all plants standing upright) to 5 (all plants prostrate) scale. Each test was harvested at maturity, and the weight and moisture content for each plot was recorded. Each plot was adjusted to 13% moisture, and the weight was converted to kg ha⁻¹.

**Gas Chromatography**

Measurements for palmitic acid, stearic acid, oleic acid, linoleic acid, and linolenic acid were done using gas chromatography for the 2009 field tests in a procedure outlined by Spencer et al. (2003). Briefly, five seeds from each plot were crushed and placed in a test tube. A 2.5 mL mixture of chloroform, hexane, and methanol (8:5:2 v/v/v) was added to each sample. The test tubes were closed with a stopper and left to sit for at least four hours. Following extraction, 100 µl of the oil sample was placed in a 1.5 mL autosampler vial. Then, 75 µl of methylation reagent [sodium methoxid:methanol:petroleum ether:ethyl ether (1:4:2 v/v/v)] and 0.75 mL of hexane were added to each vial before capping. The fatty acid compositions were determined by a Hewlett Packard HP 6890 series gas chromatograph (Palo Alto, CA) system set using a model 7683 auto sampler and a model 7673 flame ionization detector, and an immobilized 30 mm x 0.53 mm inner diameter, Alltech AT-Silar capillary column with 0.5 µm fused stationery phase. Analysis was conducted under the following operating conditions: carrier, Helium (20 ml/min); 20:1(v/v) split injection; injection temperature 250 °C; detector temperature 275 °C; and column temperature 230 °C. The RM-1 standard (appropriate for measuring soybean oil) was used in order to calibrate and determine the relative fatty acid content of the experimental samples.

**Near Infrared Reflectance Spectroscopy**

Following harvest from the 2009 growing season, approximately 25 g of seed from each plot were ground for 20 sec in a Knifetec 1095 Sample Mill (FOSS Tecator, Hoganas, Sweden). This produced whole ground soybean with a uniform consistency and particle size. Prior to analysis, the Near-Infrared Reflectance (NIR 6500, FOSS North America) instrument lamp was turned on and allowed to warm up for two hours before diagnostics were performed. The lamp was left on during the entire period of analysis, and routine diagnostics were performed each day before samples were analyzed to ensure proper instrument response, wavelength and bandwidth accuracy, and NIR repeatability. Throughout the analysis the temperature was kept at approximately 20°C and the humidity at approximately 40%. The ground samples were scanned with the NIR instrument using ISIscan software v. 2.85. This produced sample estimates for the
soybean isoflavones genistein, daidzein, and glycine, recorded in mg g⁻¹ seed weight predicted on a dry weight basis. Values for total isoflavones were obtained by summing genistein, daidzein, and glycine. The equation used for NIR prediction of genistein, daidzein, and glycine was derived from HPLC analysis using 497, 499, and 492 samples with R² values of 0.85, 0.85, and 0.65, respectively. Additionally, values for protein percentage and oil percentage were obtained on a 13% moisture basis.

**DNA Extraction and Molecular Analysis**

DNA samples were collected from crushed leaves of RILs and parents using the Qiagen Plant DNeasy Extraction Kit (Qiagen, Valencia, CA). The RILs and parents were screened at USDA-ARS, Beltsville, MD with 1,536 single nucleotide polymorphism (SNP) markers from the Universal Soybean Linkage Panel (U.S.L.P. 1.0) developed by Hyten et al. (2010). The SNPs from the U.S.L.P. 1.0 were screened using the GoldenGate assay, which was performed according to manufacturer’s protocol as described by Fan et al. (2003) and Hyten et al. (2008). Polymorphisms were detected in 480 of the SNP markers.

**Data Analysis**

The data combined over replications and locations for genistein, daidzein, glycine, and total isoflavones, for each maturity test individually, were tested for differences among RILs using the PROC MIXED procedure in SAS 9.2 (SAS Institute Inc., 2008, Cary, NC). Random blocking factors in the model included location, genotype, G × E, and replication. The PROC CORR procedure was used to obtain phenotypic correlations between genistein, daidzein, glycine, total isoflavones and other important agronomic and seed quality traits for soybean. Heritability estimates for genistein, daidzein, glycine, and total isoflavones were calculated on an entry means basis for three replications and three locations according to Nyquist (1991) as follows:

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

Where \(h^2\) represents an approximation of the narrow sense heritability, \(\sigma_g^2\) is genetic variance which is primarily additive variance in this F5 derived RIL population, \(\sigma_{ge}^2\) is genotype by environment variance, \(\sigma^2\) is error variance, \(r\) is number of replications, and \(e\) is the number of

A genetic map was estimated from this population using R/QTL package (Broman et al., 2003) in the R language and environment for statistical computing (R Development Core Team, 2009). This map, along with the least squares means (LSMEANs) combined over locations for genistein, daidzein, glycitein, and total isoflavones were used for QTL identification with QTL Cartographer software v. 2.5 (Wang et al., 2011) with the composite interval mapping (CIM) procedure. Since this was an F5 derived population, heterozygote marker loci were excluded from analyses as we were primarily interested in additive genetic effects. Analyses were conducted with the standard model Zmapqtl 6 in the CIM procedure with a 10cM window and a 1cM walk speed. The empirical likelihood of odds (LOD) thresholds was determined at the 5% level of probability with 1000 permutations for each trait in each maturity test (Churchill and Doerge, 1994). Any QTL whose LOD score exceeded the empirical LOD threshold was considered significant. A 1-LOD support interval was determined to establish a SNP marker confidence interval for QTL. Each maturity group (early, mid, and late) was tested separately for isoflavone QTL.

Results and Discussion

Phenotypic Traits

There was a significant difference among RILs for genistein, daidzein, glycitein, and total isoflavones in the early, mid, and late tests (P<0.0001). In each test, ‘Essex’ had a higher value for genistein, daidzein, and total isoflavones, while ‘Williams 82’ had a higher value for glycitein. Each test displayed a normal distribution for genistein, daidzein, glycitein, and total isoflavones (Figures 2.1 and 2.2). For each isoflavone in the early test and mid test, and for glycitein in the late test, transgressive segregation was observed, in which a large number of individual RIL values were higher or lower than either parent. This suggests that both parents carry alleles governing isoflavones heritable in progeny. Interestingly, in every test, daidzein was the most abundant isoflavone, followed by genistein and glycitein. This differs from previous studies, in which genistein is the most abundant isoflavone (Eldridge and Kwolek, 1983; Wang and Murphy, 1994; Brouns, 2002). The mid test had the highest overall mean for total isoflavone content, while the late test had the lowest. Heritability estimates were highest for the early test, with values of 0.81, 0.81, 0.54, and 0.81 for genistein, daidzein, glycitein, and total
isoflavones, respectively (Table 2.1), which were similar to the values obtained by Gutierrez-Gonzalez et al. (2009; 2010b; 2011). The mid test had similar, but slightly lower values for heritability than the early test, with values of 0.68, 0.72, 0.36, and 0.71 for genistein, daidzein, glycitein, and total isoflavones, respectively (Table 2.1). The late test had heritability values of 0.46, 0.39, 0.48, and 0.43 for genistein, daidzein, glycitein, and total isoflavones, respectively (Table 2.1), which were similar to the values obtained by Primomo et al. (2005) and Zeng et al. (2009).

Additional descriptive statistics for genistein, daidzein, glycitein, and total isoflavones are summarized in Table 2.1. The highest maximum values for daidzein, genistein, and total isoflavones were observed in the early test. While daidzein had an overall higher mean value than genistein in the early test, the maximum values were equivalent. The range for genistein, daidzein, and total isoflavones was lower and the overall mean was higher in the mid test than in the early or the late test. The late test had the lowest overall minimum value and the lowest overall mean for genistein, daidzein, and total isoflavones. The parental values were close in the early test, with both ‘Essex’ and ‘Williams 82’ exceeding the mean value for genistein, daidzein, and total isoflavones. The difference between parental values for each isoflavone except glycitein was greater in the mid test than in the early test, with ‘Essex’ above mean values and ‘Williams 82’ below mean values. The parental values for genistein, daidzein, and total isoflavones were lowest in the late test. ‘Essex’ still exceeded the mean for these traits, while ‘Williams 82’ displayed values much lower than the mean and near the minimum. Glycitein displayed a narrower range than the other isoflavones, and the parental values were close to the mean in each test.

Correlations between isoflavones and agronomic and seed quality traits are summarized in Table 2.2. Strong positive correlations (R ≥ 0.95, P < 0.0001) were found between each of the soybean isoflavones with each other, with the exception of glycitein. Genistein, daidzein, and total isoflavones displayed low to moderate positive correlations with maturity, yield, palmitic acid, linoleic acid, and linolenic acid (R ≥ 0.24, P < 0.0001). Weak negative correlations were observed between genistein, daidzein, and total isoflavones with lodging and height (R ≤ -0.17, P ≤ 0.005), but moderate negative correlations with protein, oil, stearic acid, and oleic acid (R ≤ -0.37, P < 0.0001). The significant positive correlations with yield and significant negative correlations with oil for genistein, daidzein, and total isoflavones were consistent with the results
from Primomo et al. (2005). Correlations between glycitein and other traits were either weak or insignificant at the 5% level of probability, with the exception of maturity, which was moderately negatively correlated \((R = -0.50, P < 0.0001)\), and oil, which was weakly positively correlated \((R = 0.30, P < 0.0001)\).

**Genetic Mapping and QTL Detection**

The 480 polymorphic SNP markers used to create a genetic map for this population spanned across 21 linkage groups (Figure 2.3). The map covered a total distance of 3035.4 cM, with an average distance of 6.32 cM between markers. While the overall map distance was greater than the Consensus Map 4.0 (Hyten et al., 2010), the SNP order and linkage group assignments were similar. The exceptions to this similarity were that for this study, chromosome 13 was split into two linkage groups (designated 13a and 13b), and one marker, BARC-015435-01966, was mapped to chromosome 13 rather than chromosome 6.

In total, 7 QTL for genistein (Table 2.3), 5 QTL for daidzein (Table 2.4), 3 QTL for glycine (Table 2.5), and 6 QTL for total isoflavones (Table 2.6) were detected. The QTL for each isoflavone detected on chromosome 6 and the QTL for genistein, daidzein, and total isoflavones detected on chromosome 19 could have been affected by the E1 maturity locus (chromosome 6) and the Dt1 growth habit locus (chromosome 19), respectively, both of which segregate in this population. However, previous studies have reported QTL for genistein, glycine, daidzein, and total isoflavones on chromosome 6 (Primomo et al., 2005; Zeng et al., 2009; Gutierrez-Gonzalez et al., 2011) and chromosome 19 (Gutierrez-Gonzalez et al., 2009; Gutierrez-Gonzalez et al., 2010b) in similar genetic regions to the E1 locus and the Dt1 locus, respectively. Furthermore, soybean isoflavones have been previously associated with maturity and plant height (Primomo et al., 2005). With these considerations, the QTL detected for genistein, daidzein, glycine, and total isoflavones on chromosomes 6 and 19 will be considered real effects in this study.

The QTL for genistein, designated GEN1 through GEN 7 (Table 2.3), were located on chromosomes 5, 6, 9, 9, 13b, 17, and 19 (Figure 2.4). In general, QTL detected on the same chromosome in different maturity tests were considered to be the same if the additive effect was the same and they were located within 10 cM of each other. However, the QTL detected on chromosome 9 in the early test was considered to be different from the QTL on chromosome 9 in the late test, because the distance between them exceeded 40 cM (Figure 2.4). The \(R^2\) values for
QTL detected for genistein in this study ranged from 0.09-0.23 (Table 2.3), representing large genetic effects. Four of the QTL detected for genistein in this study, located on chromosomes 5, 9, 9, and 17, have not been previously reported. Of these, the QTL on chromosome 9 detected in the early test had the largest $R^2$ value (0.23). This finding represents a putative new major QTL for genistein which could be useful for MAS. Following examination of Consensus Map 4.0 (Hyten et al., 2010), it was determined that the QTL for genistein in similar genetic regions on chromosomes 6 (Primomo et al., 2005; Zeng et al., 2009; Gutierrez-Gonzalez et al., 2011), 13 (Zeng et al., 2009; Gutierrez-Gonzalez et al., 2009; Gutierrez-Gonzalez et al., 2010b), and 19 (Gutierrez-Gonzalez et al., 2009; Gutierrez-Gonzalez et al., 2010b) have been previously reported. These findings are interesting for several reasons. The studies conducted by Primomo et al. (2005) and Zeng et al. (2009) were conducted in extremely different environments using different parents from this study, and so finding the same QTL under these conditions represents a validation consistency. The studies by Gutierrez-Gonzalez et al. (2009; 2010b) as well as this study, all used ‘Essex’ as a parent in population development. However, in this study, ‘Essex’ was the high isoflavone parent, whereas in Gutierrez-Gonzalez et al. (2009; 2010b), ‘Essex’ was the low isoflavone parent. Additionally, in Gutierrez-Gonzalez et al. (2009; 2010b), ‘Essex’ was the low parent for the genistein QTL on chromosome 13 and the high parent on chromosome 19. The same effects for ‘Essex’ were observed in this study for the genistein QTL detected on chromosomes 13b and 19. The re-occurrence of these QTL in different environments, but with a common parent, further demonstrates a consistency that is important when considering heritable QTL for MAS.

Based on our results for a fully additive genetic model, MAS assembling all QTL in the allelic form for higher genistein in each maturity test could potentially increase that isoflavone by 19%, 9%, and 7% above the mean value in the early, mid, and late tests, respectively. This would provide a targeted prediction of 0.82 mg g$^{-1}$ and 0.73 mg g$^{-1}$ genistein for the early and late tests, both of which exceed the values of the high parent (‘Essex’) in their respective tests. The predicted value for MAS in the mid test is 0.82 mg g$^{-1}$, which is below the value for ‘Essex’ (0.84 mg g$^{-1}$). For the early, mid and late tests, the predicted improvement for MAS of genistein is below the maximum value, suggesting that additional genistein QTL undetected by this study may segregate in this population.
The QTL detected for daidzein, designated as DAI1 through DAI5 (Table 2.4) in this study were located on chromosomes 5, 6, 9, 13, and 19 (Figure 2.5). All were major QTL with $R^2$ values ranging from 0.10-0.30. It should be noted that both the lowest (0.10) and the highest (0.30) $R^2$ values detected for daidzein were for the same QTL detected in different maturity tests. This difference helps to illustrate the influence of maturity on genetic control of isoflavone content. Of the QTL detected for daidzein, only the one located on chromosome 5 had not been previously reported. QTL in similar genetic regions to those detected in this study on chromosomes 6 (Gutierrez-Gonzalez et al., 2011), 9 (Kassem et al., 2004), 13 (Primomo et al., 2005; Zeng et al., 2009), and 19 (Gutierrez-Gonzalez et al., 2009) had all been identified by previous research efforts and are validated in this study. Moreover, Gutierrez-Gonzalez et al. (2009) and Kassem et al. (2004) also used ‘Essex’ as a parent in population development. As in this study, Essex was the high parent for the QTL detected on chromosome 9 (Kassem et al., 2004) and chromosome 19 (Gutierrez-Gonzalez et al., 2009).

Using our results for a fully additive genetic model, an improvement of 13%, 4%, and 2% for daidzein over the mean value in the early, mid, and late tests, respectively, could be observed. This would provide predicted values of 0.90 mg g$^{-1}$, 0.87 mg g$^{-1}$, and 0.80 mg g$^{-1}$ daidzein in the early, mid, and late tests, respectively; each greater than the high parent (‘Essex’) value, but less than the maximum value for daidzein in their respective tests, suggesting that additional daidzein QTL that were not detected in this study are segregating in this population.

The QTL detected for glycine, designated as GLY1 through GLY3 (Table 2.5), were located on chromosomes 6, 9, and 20 (Figure 2.6). The $R^2$ values for glycine QTL ranged from 0.15-0.20, indicative of major QTL. Fewer QTL were detected for glycine than for any of the other isoflavones, with none being detected in the late maturity test. This may be the result of the lower heritability values for glycine in relation to the other isoflavones. QTL in similar genetic regions to those detected in this study on chromosomes 6 (Gutierrez-Gonzalez et al., 2010b), and chromosome 9 (Gutierrez-Gonzalez et al., 2009) have been previously reported in studies containing ‘Essex’ as a parent. Gutierrez-Gonzalez et al., (2010b) reported ‘Essex’ as the low parent for the glycine QTL on chromosome 6, which was consistent with the results from this study. However, Gutierrez-Gonzalez et al. (2009) found ‘Essex’ to the low parent for chromosome 9, which was different from this study in, as ‘Essex’ was the high parent. QTL in
similar genetic regions to those detected in this study on chromosome 20 has been previously reported by Zeng et al. (2009) in which different parental lines were used.

Using a fully additive genetic model to estimate MAS improvements, an increase of 1% in both the early and the mid tests, and 0% in the late test, as no glycitein QTL were detected in the late test. These minimal improvements over the mean are still below the high parent (‘Williams 82’) values in each test. The use of MAS for glycitein in this population would likely not be worth the effort, and phenotypic selection would be preferable.

Since each of the isoflavones has beneficial effects for humans, the identification of QTL for total isoflavone content is perhaps the most important information to obtain for use in genetic improvement. The QTL detected for total isoflavone content in this study, designated as ISO1 through ISO6 (Table 2.6), were located on chromosomes 5, 5, 6, 9, 13, and 19 (Figure 2.7). As with daidzein, which was a large contributor for total isoflavones, 4 QTL for total isoflavones were detected in the early test, while only 2 QTL were detected in the mid and late tests. The heritability values were once again much lower in the late test (0.43) than in the early (0.81) or the mid (0.71) tests. For the QTL detected for total isoflavone content, the $R^2$ values ranged from 0.09-0.21. Both of the QTL on chromosome 5 and the QTL on chromosome 9 have not been previously reported. Of these, the QTL on chromosome 9 is notable as it explained the greatest variation in total isoflavone content among the QTL detected. QTL in similar genetic regions to those detected in this study on chromosomes 6 (Primomo et al., 2005; Gutierrez-Gonzalez et al., 2011), 13 (Zeng et al., 2009), and 19 (Gutierrez-Gonzalez et al., 2009; Gutierrez-Gonzalez 2010b) had been detected previously. This study, as well as Gutierrez-Gonzalez et al. (2009; 2010b), used ‘Essex’ as a parent. For the QTL detected on chromosome 19 by Gutierrez-Gonzalez et al. (2009; 2010b) ‘Essex’ was the high parent, which was consistent with the results from this study.

A fully additive genetic model estimating the effects of MAS in this population would predict values exceeding the mean by 15%, 4%, and 2%, in the early, mid, and late tests, respectively. These improvements would represent values of 1.91 mg g$^{-1}$, 1.82 mg g$^{-1}$, and 1.66 mg g$^{-1}$ for the early, mid, and late tests, respectively. These potential improvements using MAS are greater than the high parent (‘Essex’) value in the early and late tests, but less than ‘Essex’ in the mid test. The early test had a higher heritability value for total isoflavone content and more QTL detected than either the early or the late test, so it is consistent with expectations that the
early test would exhibit the greatest potential for improvement through MAS. However, in each test, the maximum value for total isoflavone content exceeded the predicted MAS value, indicating that more total isoflavone QTL segregate in this population than were detected.

**Conclusion**

The correlations between genistein, daidzein, and total isoflavones with other important agronomic and seed quality traits were somewhat troubling with the exception of a positive correlation with yield and a negative correlation with lodging. For soybean, reduction of protein content, oil content, and oleic acid content are undesirable, while reduction in maturity, palmitic acid, and linolenic acid are often targeted goals of improvement. Future studies should continue research to better understand the relationships between soybean isoflavones and other important agronomic and seed quality traits as breeders strive to optimize improvements.

Overall, 21 QTL for genistein, daidzein, glycitein, or total isoflavones were detected by this study, of which 8 are newly detected and 13 have been previously identified. The maturity tests showed some variability in the QTL which were detected for each trait. In general, QTL detected in both the early and the mid test had a higher LOD score and a higher $R^2$ value than QTL in similar genetic regions detected in the late test. This result, combined with the higher heritability values and more QTLs detected in the early test than the mid or late tests lead to the potential for greater genetic improvement for MAS in the early test. The lower number of total QTL observed in the late test could have resulted from the lower heritability values observed in comparison to the early test and mid test. For each trait, the differences in the number of QTL detected by maturity test, as well as the differing $R^2$ values and additive effects for similar QTL detected in different maturity tests help to illustrate the effect that soybean maturity can have on isoflavone content (Primomo et al., 2005).

Many of the QTL detected in this study for different isoflavones were in similar genetic regions to one another. For example, five of the QTL detected for genistein (GEN1, GEN2, GEN3, GEN4, and GEN6, respectively) were in very similar genetic regions to QTL detected for daidzein (DAI1, DAI2, DAI3, DAI5, and DAI4, respectively) and for total isoflavones (ISO1, ISO2, ISO3, ISO5, and ISO4, respectively). As genistein and daidzein are the primary components of total isoflavones in soybean, it seems evident that the QTL for genistein or daidzein would be the same as those detected for total isoflavones. For comparison of the QTL detected for genistein, daidzein, and glycitein, examination of the phenylpropanoid pathway
from which they are synthesized helps to shed some light on the results (Figure 2.8). Synthesis of each of the isoflavones share many steps in common, and so QTL detected in similar genetic regions for different QTL could possibly represent genes controlling enzyme expression for common steps in the phenylpropanoid pathway. Glycitein has a different branch point than daidzein and genistein, and so it has fewer enzymes in common with the other isoflavones. This could explain some of the differences in position for glycine QTL in comparison with QTL for other isoflavones.

The QTL detected in this study are important for several reasons. The parents chosen for this study both have significance, as ‘Essex’ has been used for several previous isoflavone detection studies (Meksem et al., 2001; Kassem et al., 2004; Kassem et al., 2006; Gutierrez-Gonzalez et al., 2009; Gutierrez-Gonzalez et al., 2010b), and ‘Williams 82’ has been used in the sequence of the soybean genome (Schmutz et al., 2010). In addition, ‘Essex’ is a prominent ancestor of modern southern U.S. cultivars, and ‘Williams’, from which ‘Williams 82’ was derived, is a prominent ancestor of modern northern U.S. cultivars (Hyten et al., 2004). The addition of 8 new QTL for genistein, daidzein, glycine, or total isoflavones adds to the understanding of the genetic regions controlling isoflavone expression. Of particular interest among these 9 new QTL are those for genistein and for total isoflavone content detected on chromosome 6 in the early test, which had relatively high $R^2$ values in comparison to the other QTL detected for those traits. The detection of 13 previously reported QTL for genistein, daidzein, glycine, and total isoflavones grown in different environments, using different as well as similar genetic material is noteworthy in identifying consistent genetic regions controlling isoflavone content. The importance of validation for these 13 QTL is further illustrated in that most previous studies used HPLC analysis for isoflavone detection, while this study used the much faster and lower cost method of NIR spectroscopy. Furthermore, this study used a large number of polymorphic markers throughout the soybean genome, which is important for QTL detection (Yang et al., 2011). The QTL detected by this study could be beneficial for MAS, resulting in the genetic improvement of soybean isoflavones.
References
References


CHAPTER 3

COMPARISON OF MAS AND PHENOTYPIC SELECTION METHODS
FOR SOYBEAN ISOFLAVONE GENISTEIN
Abstract

In recent decades, many QTL detection studies with the potential for use in marker-assisted selections (MAS) have been published. However, few studies exist testing the effectiveness of MAS. Evidence for the benefits of MAS for qualitative traits is strong, but more work is needed for testing the effectiveness of MAS for quantitative traits. The implementation of MAS for genistein, which is a quantitative trait, would be useful as phenotyping data is costly and time consuming. This study sought to test of effectiveness of MAS for the soybean [Glycine max (L.) Merrill] isoflavone genistein in comparison with phenotypic selections. The comparison of selection methods was performed in field tests using four different soybean maturity groups (III Late, IV Early, IV Late, and V Early) grown in 2010 and 2011 in an RCBD with 3 reps and 3 locations (Knoxville, TN; Springfield, TN; Milan, TN). Overall, the phenotypic selections outperformed the marker-selections for genistein; however, due to an error which occurred in the initial marker-selections, not as many RILs were available for comparison as was originally intended. Genistein was positively correlated with the other soybean isoflavones, as well as with yield. The correlation with yield, as well as potential health benefits derived from genistein observed in previous studies, reflect the need for genetic improvement of isoflavone expression. Future studies exploring the use of MAS for improved genistein are warranted.

Introduction

Traditional plant breeding has used analytical measurements to make selections for genetic improvement. An example of this is yield improvement for soybean, which rose at a linear rate of 22.6 kg ha\(^{-1}\) during the period from 1924 to 1998 (Specht et al., 1999), largely due to phenotypic breeding. However, for many traits this method can be time consuming and costly. As molecular markers have recently become more available and more cost effective, increased efforts have focused on marker-assisted selection (MAS). Using molecular markers, selections can be made earlier in the breeding process so that less material needs to be screened and advanced. This results in saved time and resources in comparison with phenotypic selections (Torres et al., 2010). With MAS, selections can be made at offsite nurseries representing a low heritability environment (Bernardo, 2008), which is not practical with
phenotypic selection. MAS is also beneficial for traits which are costly and time consuming to phenotype (Babu et al., 2004). Significant advantages of MAS in comparison with phenotypic breeding are exhibited with backcross breeding, in which a greater percentage of the recurrent parent can be retained in less time when incorporating traits of interest from the donor parent (Babu et al., 2004; Xu and Crouch, 2008; Pantalone et al., 2010). Considering the potential benefits of MAS as a selection method, several studies have sought to determine its effectiveness. Concibido et al. (1996) used MAS to select for soybean cyst nematode (Heterodera glycines Ichinohe; SCN) resistance and determined that it was similar in accuracy but was more time and cost effective than conventional screening techniques. Walker et al. (2002; 2004) used MAS to incorporate insect resistance into soybean lines. Maroof et al. (2008) used MAS to pyramid three soybean mosaic virus resistant genes into soybean lines. Further, Scaboo et al. (2009) used MAS to incorporate low phytate quantitative trait loci (QTL) into soybean. The usefulness of MAS appears evident for traits controlled by few loci.

More challenges arise when considering MAS for quantitative traits, including genotype x environment interactions and epistatic interactions (Babu et al., 2004; Xu and Crouch, 2008). However, such challenges can be overcome with the aid of molecular technology advances, improved experimental design, and improved statistical procedures (Babu et al., 2004; Xu and Crouch, 2008). An important consideration when selecting QTL for use in MAS for quantitative traits includes QTL detection in high heritability environments (Bernardo, 2008). Further, QTL used in MAS should also be detected using a large number of evenly spaced markers (Xu and Crouch, 2008; Yang et al., 2011) and with a large population (Bernardo, 2008). Improvements in high-throughput low cost genotyping and accurate phenotyping will also improve MAS potential in quantitative traits (Babu et al., 2004).

Because of important potential health benefits derived from soybean isoflavones, including possible cancer prevention (Birt et al., 2001), reduced symptoms from diabetes and obesity (Bhathena and Velasquez, 2002), reduced risk of coronary heart disease (Sagara et al., 2003), and reduced menopausal symptoms (Brouns, 2002), several studies have sought to detect QTL controlling soybean isoflavone expression (Meksem et al., 2001; Primomo et al., 2005; Gutierrez-Gonzalez et al., 2011). However, at this time, no studies have attempted to assess the usefulness of MAS for soybean isoflavones in comparison with phenotypic selections. A
possible explanation for this could be the perceived difficulty of using MAS for a trait that is controlled by many genes with a large complexity of epistatic interactions (Gutierrez-Gonzalez et al. 2010).

Using MAS for isoflavone improvement could be important, as lab techniques to measure soybean isoflavones can be prohibitively time consuming and costly. Encouragement for the usefulness of MAS for quantitative traits is illustrated by Sebastian et al. (2010), in which MAS was used for yield improvement in soybean. Similar to soybean isoflavones, yield is a highly complex trait influenced by both genotype and environment (Sebastian et al., 2010).

As genistein is typically the most abundant soybean isoflavone (Brouns, 2002), and is strongly correlated with daidzein and total isoflavone content (Primomo et al., 2005), it would be useful to test the effectiveness of MAS for genistein in comparison with phenotypic selections. Considering this, the objectives of this current research were to (i) use QTL for genistein to make marker-assisted selections; (ii) compare marker-assisted selections with phenotypic selections for genistein; and (iii) examine phenotypic correlations between genistein and other important agronomic and seed quality traits.

However, due to an error which occurred at the beginning of this research, marker-assisted selections were chosen for plant height rather that for genistein in many cases. Because of this, comparison of high and low marker-assisted selections for plant height became a secondary consideration of this research, and those results are also reported.

**Material and Methods**

**Plant Materials**

A population of 274 RILs was developed from the parental lines ‘Essex’ and ‘Williams 82’. The ‘Essex’ and ‘Williams 82’ cultivars used in this population were obtained from the USDA soybean germplasm collection (www.ars-grin.gov), and a random single plant was intentionally selfed for two generations to provide highly homozygous parental lines to be crossed for RIL development. The initial cross was made in the summer of 2005 at the East Tennessee Research and Education Center (ETREC) in Knoxville, TN. The F<sub>1</sub> single plants were harvested in the fall of 2005 and grown as F<sub>1</sub> single plants in Puerto Rico at the Tropical Agricultural Research Station (TARS) in Isabela, Puerto Rico, in the winter of 2005-06. Following the single seed descent method (Brim, 1966), this population was advanced from the
F₂ to the F₅ generation as follows: the F₂ seed harvested from TARS were grown at ETREC in the summer of 2006 and F₂:3 seed were harvested in the fall of 2006; the F₂:3 seed were grown at ETREC in the summer of 2007 and F₃:4 seed were harvested in the fall of 2007; the F₃:4 seed were grown at TARS in the winter of 2007-2008, F₄:5 seed were harvested at maturity and grown at TARS in the Spring of 2008; the F₅:6 plants harvested from TARS were grown in Knoxville in the summer of 2008 as RILs in 3.1 m rows for agronomic data collection and leaf collection for DNA extraction. The F₅:7 RILs harvested in the fall of 2008 were sent to Homestead, FL, for seed increase in the winter of 2008-2009. Seed harvested from the spring of 2009 were grown as 274 F₅:8 RILs in the summer of 2009 tested in replicated yield trials. The 274 RILs were divided into three different tests based on maturity from the 2008 data (early, mid, and late), with each test containing 91 or 92 RILs. Both parents, along with four other check cultivars or lines were included in each test. The check lines for the early test (‘IA4004’, LD00-2817P, LD00-3309, and ‘Macon’), mid test (TN05-4008, TN06-189, TN06-196, and ‘5002T’), and late test (JTN-5203, ‘Osage’, ‘5002T’, and ‘5601T’), were selected to correspond with the appropriate maturity group for their respective test. The 2009 field trials were grown using a randomized complete block design (RCBD) replicated three times and grown in three different locations (Knoxville, TN; Harrisburg, IL; and Stuttgart, AR). Each entry was planted in a plot consisting of two adjacent rows 6.1 m in length, with the rows spaced 0.8 m apart.

Maturity data from the 2009 field tests was used to assign four maturity groups (MGs) based on a linear regression line (R² = 0.89) for the 2010 field tests: 3 late (3L), 4 early (4e), 4 late (4L), and 5 early (5e). Phenotypic selections at 2.5% intensity for genistein were made from both tails for each of the four maturity groups, resulting in 2, 2, 6, and 4 RILs in the 3L, 4e, 4L, and 5e MG tests, respectively. Two additional lines consisting of either a check or a parental line were included in each test. The additional cultivars for the MG 3L test (‘Williams 82’ and IA4004) MG 4e test (IA4004 and LD00-2817P) MG 4L test (LD00-2817P and 5002T) and MG 5e test (5002T and ‘Essex’) were chosen to correspond with appropriate maturity. In total, there were four different phenotypic field tests, designated as Gen3L, Gen4e, Gen4L, and Gen5e.

Three separate field tests were developed using MAS. A selection intensity of 2.5% was imposed on RILs with high or low QTLs for genistein within each of the MGs. The number of random selections among those RILs was made so that the number of high and low RILs selected
for each MG were equivalent to the corresponding phenotypic test: 2, 6, and 4 for the 4e, 4L, and 5e MG tests, respectively. There was not a MAS field test for the 3L MG because no high tail selections were originally detected. As with the phenotypic tests, two additional lines consisting of either a check or a parental line were included in each MAS test. The additional cultivars are the same as those described above for the MG 4e test, MG 4L test, and MG 5e test. The three tests developed using MAS were designated as Gen4eMAS, Gen4LMAS, and Gen5eMAS.

The 2010 field tests were repeated in 2011 with some new additions to the MAS tests. New RILs were added to the 4LMAS test (7 high selections and 1 low selection) and the 5eMAS test (7 high selections). The 2010 and 2011 field tests were planted in a RCBD replicated three times and grown in three different locations (Knoxville, TN; Springfield, TN; and Milan, TN) representing differing geographic regions of East, Middle, and West TN, respectively. Field tests in the same maturity group, e.g. Gen4L and Gen4LMAS, were planted adjacently at each location. Each test entry was planted in a plot consisting of two adjacent rows 6.1 m in length, later end-trimmed to 4.9 m, with the rows spaced 0.8 m apart. The exceptions to this were the 2010 field tests at Milan, which remained at 6.1 m throughout the field season.

Agronomic Traits

For the 2009, 2010, and 2011 growing seasons, flower color was recorded at the R2 growth stage when 95% of the plants in a plot were in full bloom. Pubescence color and date of maturity were recorded during the R8 growth stage when 95% of the pods in a plot showed their mature color (Fehr and Caviness, 1977). Also at the R8 growth stage, height measurements were taken and lodging estimates were recorded. Height for each plot was estimated to be the average height of the plants within the plot. Lodging estimates were taken on a 1 (all plants standing upright) to 5 (all plants prostrate) scale. Each test was harvested at maturity, and the weight and moisture content for each plot was recorded. Each plot was adjusted to 13% moisture, and the weight was converted to kg ha$^{-1}$.

Gas Chromatography

Measurements for palmitic acid, stearic acid, oleic acid, linoleic acid, and linolenic acid were done using gas chromatography for the 2009, 2010, and 2011 field tests in a procedure outlined by Spencer et al. (2003). Briefly, five seeds from each plot were crushed and placed in a test tube. A 2.5 mL mixture of chloroform, hexane, and methanol (8:5:2 v/v/v) was added to
each sample. The test tubes were closed with a stopper and left to sit for at least four hours. Following extraction, 100 µl of the oil sample was placed in a 1.5 mL autosampler vial. Then, 75 µl of methylation reagent [sodium methoxid:methanol:petroleum ether:ethyl ether (1:4:2 v/v/v)] and 0.75 mL of hexane were added to each vial before capping. The fatty acid compositions were determined by a Hewlett Packard HP 6890 series gas chromatograph (Palo Alto, CA) system set using a model 7683 auto sampler and a model 7673 flame ionization detector, and an immobilized 30 mm x 0.53 mm inner diameter, Alltech AT-Silar capillary column with 0.5 µm fused stationery phase. Analysis was conducted under the following operating conditions: carrier, Helium (20 ml/min); 20:1(v/v) split injection; injection temperature 250 °C; detector temperature 275 °C; and column temperature 230 °C. The RM-1 standard (appropriate for measuring soybean oil) was used in order to calibrate and determine the relative fatty acid content of the experimental samples.

**Near Infrared Reflectance Spectroscopy**

Following harvest from the 2009, 2010, and 2011 growing seasons, approximately 25 g of seed from each plot were ground for 20 sec in a Knifetec 1095 Sample Mill (FOSS Tecator, Hoganas, Sweden). This produced whole ground soybean with a uniform consistency and particle size. Prior to analysis, the Near-Infrared Reflectance (NIR 6500, FOSS North America) instrument lamp was turned on and allowed to warm up for two hours before diagnostics were performed. The lamp was left on during the entire period of analysis, and routine diagnostics were performed each day before samples were analyzed to ensure proper instrument response, wavelength and bandwidth accuracy, and NIR repeatability. Throughout the analysis the temperature was kept at approximately 20°C and the humidity at approximately 40%. The ground samples were scanned with the NIR instrument using ISIscan software v. 2.85. This produced sample estimates for the soybean isoflavones genistein, daidzein, and glycitein, recorded in mg g⁻¹ seed weight predicted on a dry weight basis. Values for total isoflavones were obtained by summing genistein, daidzein, and glycitein. The equation used for NIR prediction of genistein, daidzein, and glycitein was derived from HPLC analysis using 497, 499, and 492 samples with R² values of 0.85, 0.85, and 0.65, respectively. Additionally, values for protein percentage and oil percentage were obtained on a 13% moisture basis.
DNA Extraction and Molecular Analysis

DNA samples were collected from crushed leaves of RILs and parents using the Qiagen Plant DNeasy Extraction Kit (Qiagen, Valencia, CA). The RILs and parents were screened with 1,536 single nucleotide polymorphism (SNP) markers from the Universal Soybean Linkage Panel (U.S.L.P. 1.0) developed by Hyten et al. (2010). The SNPs from the U.S.L.P. 1.0 were screened using the GoldenGate assay, which was performed according to manufacturer’s protocol as described by Fan et al. (2003) and Hyten et al. (2008). Polymorphisms were detected in 480 of the SNP markers.

QTL Detection and Marker-Assisted Selections

A genetic map was estimated from this population using R/QTL package (Broman et al., 2003) in the R language and environment for statistical computing (R Development Core Team, 2009). This map, along with the least squares means (LSMEANs) combined over locations for genistein and plant height from the 2009 field tests were used for QTL identification with QTL Cartographer software v. 2.5 (Wang et al., 2011) with the composite interval mapping (CIM) procedure. Since this was an F5 derived population, heterozygote marker loci were excluded from analyses as we were primarily interested in additive genetic effects. Analyses were conducted with the standard model Zmapqtl 6 in the CIM procedure with a 10cM window and a 1cM walk speed. The empirical likelihood of odds (LOD) thresholds were determined at the 5% level of probability with 1000 permutations for each trait in each maturity test (Churchill and Doerge, 1994). Any QTL whose LOD score exceeded the empirical LOD threshold was considered significant. Each 2009 maturity group (early, mid, and late) was tested separately for QTL. A 1-LOD support interval was determined to establish a confidence interval for significant QTL. Marker-Assisted selections were made by selecting all of the SNP markers within the 1-LOD range for all QTLs in the 2009 test in which the selected RIL was grown (Figures 3.1, 3.2, 3.3, 3.4, 3.5, and 3.6). RILs containing either favorable or unfavorable alleles were considered candidates for opposite tail MAS to be compared with high or low phenotypic tail selections, respectively. Candidate RILs for MAS were randomly chosen among those exhibiting the QTL criteria for entry into the 2010 and 2011 field tests.
Data Analysis

Data for genistein and plant height from the 2010 and 2011 field tests averaged across locations and years were tested for differences among genotypes and tail selections using the PROC MIXED procedure in SAS 9.2 (SAS Institute Inc., 2008, Cary, NC). For comparison of genistein selections, each MG was analyzed together, and Tukey-Kramer mean separation was used for comparison of lines and tail selections. Since the 3L MG did not contain a MAS test, a simple RCBD model was used, in which the fixed term was genotype and the random terms were environment, rep(environment), and genotype*environment. In the 3L MG, tail was not included as a fixed term in the model, because there was only one genotype within each tail, and so genotype comparisons could also be used for tail comparisons. Even though the random term genotype*environment was part of the model, it was not included in the written PROC MIXED statement so that the correct error term would be tested by the model. The PROC MIXED procedure uses the simplest random term that contains each fixed term for testing of RCBD models, and so to avoid incorrectly testing the fixed term with the random genotype*environment term, it was excluded from the PROC MIXED statement and its degrees of freedom were pooled with the error term.

For the 4e, 4L, and 5e maturity test comparisons of genistein, an RCBD model was used with genotype(test*tail), test, and tail(test) as the fixed terms; and rep(environment), environment, and environment*genotype(test*tail) as the random terms. Once again, the random term environment*genotype(test*tail) was excluded from the written PROC MIXED statement to ensure use of the correct error term for testing of fixed terms.

Since no phenotypic selections were made for height, comparisons were only made between tail selections in the Gen4eMAS, Gen4LMAS, and Gen5eMAS tests, individually. The RCBD model used for height comparison included genotype(tail) and tail as fixed terms, and rep(environment), environment, and environment*genotype(tail) as random terms. As with the models used for genistein, the random term environment*genotype(tail) was not included in the PROC MIXED statement to ensure use of the correct error term. Tukey-Kramer mean separation was used for comparison of lines and tail selections.
The PROC CORR procedure in SAS 9.2 (SAS Institute Inc., 2008, Cary, NC) was used to examine phenotypic correlations between genistein and other important agronomic and seed quality traits for the 2010 and 2011 field seasons combined over years, consisting of 41 RILs.

**Results and Discussion**

The initial intent of this research was to compare phenotypic selections with MAS in high and low tail selections for the soybean isoflavone genistein. However, in the summer of 2011, after all of the field tests had been planted and well established, it was discovered that many of QTL initially detected were for plant height and not for genistein. This was due to a mistaken selection of height rather than genistein in the drop down menu of QTL Cartographer software v. 2.5 (Wang et al., 2011). The initial QTL detection was done in the spring of 2010, prior to the initial planting of the selection based field tests, and so a comparison between high and low marker assisted selections for plant height in the 2010 and 2011 MAS field tests was possible, but phenotypic height selections were not. There was 1 QTL detected for height (Table 3.1) on chromosome 19, and it was detected independently in each of the 2009 maturity tests. This was a major QTL with a maximum $R^2$ value of 0.77. It is likely that this QTL represents the finding of the Dt1 growth habit locus, as it was in the same genomic region as the Dt1 locus, which segregates in this population. Of the parental lines, ‘Essex’ has a determinate growth habit, while ‘Williams 82’ has an indeterminate growth habit. The favorable parent for plant height on the QTL detected in this study was ‘Williams 82’, which is consistent with the expectations from this effect.

For the comparison of height selections, the 4eMAS test contained two RILs that were high selections for plant height and two check cultivars. Using Tukey-Kramer mean separation, it was determined that the two high selections were not significantly different from one another (Table 3.2). Most of the selections for genistein were either reversed or invalidated when used as height selections. This was the case for the ExW3-203, which was originally intended as a high genistein selection, but after the error was detected, it was determined to be an invalid genistein selection and a high height selection in the 4eMAS test. However, ExW3-164, which was originally intended as a low genistein selection was valid, and the height tail exhibited the reverse of this effect as a high selection. Thus, both RILs in the 4eMAS test were high selections for height.
The 4LMAS test contained seven high selections for plant height, seven low selections for plant height, and two check cultivars. Each of the high RIL selections significantly exceeded the low RIL selections in plant height (Table 3.2). Additionally, the high selections demonstrated a greater overall mean for plant height than the low selections (Table 3.3; Figure 3.7).

For the 5eMAS test, there were nine high selections for plant height, two low selections for plant height, one check cultivar, and one parent (‘Essex’). The 2010 field test contained two high selections and two low selections; however, seven high selections were added in the 2011 test. No low selections were added as none were available at the time of planting for the 2011 field test. As with the previous tests, each of the high selections expressed significantly greater plant height than any other line in the test (Table 3.2). Once again, the mean high selection exceeded the low selections as well as the parental line (Table 3.3; Figure 3.7). ‘Essex’, which was the unfavorable parent for plant height, expressed the lowest overall height in the 5eMAS test (Table 3.2).

After the mistake of QTL detection was discovered, a corrected QTL detection was done for genistein in the summer of 2011, and it was discovered that several of the MAS tail selections were also valid for genistein comparison with the phenotypic selections. However, since the MAS tests had initially been set up with height QTL, many of the high and low tail selections for genistein were reversed, while many of the tail selections were invalid. The invalid tail selections were still left in the field tests in order to determine if MAS for genistein would be different from unselected RILs in the test.

The initial setup of the phenotypic tests was unchanged and the selections remained valid. In the 4eMAS test, no new RILs were added, and there was one low selection and one unselected RIL when the selections had been corrected. After the correction of the 4LMAS test, there were four high selections, no low selections, and ten unselected RILs for genistein. The 5eMAS test had two high selections, no low selections, and nine unselected RILs for genistein after the correction.

Analysis of the 3L maturity group from the 2010 and 2011 field tests determined that the high phenotypic selection significantly outperformed the low phenotypic selection for genistein (Table 3.4). Further, ‘Williams 82’, which was the low parent for genistein, was significantly
lower than the high selection, but ‘Williams 82’ was higher than the low selection (Table 3.5; Figure 3.8). Since there were no MAS selections in the 3L maturity group, no comparisons were made between phenotypic selections and MAS.

For the 4e maturity group, the line with the greatest genistein content was an unselected RIL from the Gen4eMAS test; however, it was not significantly different from the high phenotypic RIL (Table 3.4). Comparison of the low tail selections from the Gen4e and the Gen4eMAS tests showed that selections from phenotypic and MAS approaches were not significantly different from each other, but they were significantly lower than the high phenotypic selection and the unselected MAS RIL (Table 3.5, Figure 3.8). While not chosen as entries in the 4eMAS test, both the high and low phenotypic selections were candidates for high and low marker-selections, respectively.

The two highest entries for genistein content in the 4L maturity group were high phenotypic selections; however, they were not significantly greater than the next two RILs, which were high marker-selections (Table 3.4). The highest genistein content was achieved by a high phenotypic selection, which was also a candidate as a high selection RIL in the MAS test. Comparison of tail selections revealed that high phenotypic selections were the overall highest, followed by high MAS selections (Table 3.5, Figure 3.8). Low phenotypic selections were the lowest group for genistein content.

As with the 4L maturity group, the two highest lines in the 5e maturity group were both high phenotypic selections, which were both significantly higher in genistein content than the two high marker-selections (Table 3.6). Both high marker-selections exceeded ‘Essex’, which was the high parent, in genistein content, suggesting that MAS successfully identified transgressive segregates for genistein. The high phenotypic selections had the highest overall genistein content, while the low phenotypic selections had the lowest overall genistein content. While lower than the high phenotypic selections, the high marker-selections did contain significantly higher genistein than the unselected MAS RILs (Table 3.5, Figure 3.8).

Correlations between genistein and other important agronomic and seed quality traits are summarized in Table 3.7. Genistein had a strong positive correlation with daidzein and total isoflavones ($r \geq 0.96$, $P < 0.0001$) and a moderate positive correlation with glycine (r = 0.61, $P < 0.0001$). For agronomic traits, genistein had a weak positive correlation with maturity ($r =$
0.32, \( P < 0.05 \)), a moderate positive correlation with yield \((r = 0.67, \ P < 0.0001)\), and insignificant correlations with lodging and height at the 5% level of probability. Primomo et al. (2005) previously observed a significant positive correlation between genistein and yield. A strong negative correlation was observed between genistein and protein in this study \((r = -0.83, \ P < 0.0001)\), which has also been noted by Wilson (2004) and Primomo et al. (2005), while an insignificant correlation was observed between genistein and oil at the 5% level of probability. Genistein showed a moderate negative correlation with stearic acid \((r = -0.54, \ P < 0.001)\), a strong negative correlation with oleic acid \((r = -0.75, \ P < 0.0001)\), a moderate positive correlation with palmitic acid and linoleic acid \((r \geq 0.43, \ P < 0.01)\), and a strong positive correlation with linolenic acid \((r = 0.80, \ P < 0.0001)\) (Table 3.7).

### Conclusion

A mistake in the initial QTL detection for genistein yielded the opportunity to compare high and low marker-selections for soybean height. In the 4eMAS test, the high selections were not statistically different. In the 4LMAS and 5eMAS tests, the high selections were significantly higher than the low selections. These results show that selecting for height QTL can influence height expression. Comparison of height marker-selections with height phenotypic selections would have been useful, but this opportunity was not available as no phenotypic selections were made for height. Soybean height in this population is apparently controlled by fewer genes than genistein, as evidenced by the finding of only one large height QTL in this population. The effect of the major QTL was apparent in this population as the high marker-selections grew taller than the low marker-selections in every instance.

In contrast to height, genistein is a quantitative trait controlled by many genes (Gutierrez-Gonzalez et al., 2009; Zeng et al., 2009). While the number of RILs for MAS and phenotypic selection comparison was less than initially intended, there were still enough remaining valid selections to compare selection methods. In every instance, high phenotypic selections exceeded low phenotypic selections in genistein content. High phenotypic selections outperformed high marker-selections for genistein, while low phenotypic selections were not different from low marker-selections for genistein. The phenotypic low selection was also a low MAS candidate for genistein. While none were selected for the MAS tests, several of the high and low phenotypic selections were MAS candidates. Considering this, the MAS tests may have been more
competitive with the phenotypic tests if more MAS candidates had been chosen. Also, the intense 2.5% threshold for phenotypic genistein selections influenced their high performance. Had a more liberal threshold for phenotypic selection been chosen, the marker-selections may have been more competitive. Additionally, a better comparison would have been possible if more RILs had been selected, which would have increased the statistical power of comparison.

The significant positive correlations between genistein and the other isoflavones show that selecting for genistein could also be used to increase any of the isoflavones. Also, the significant positive correlation between genistein and yield indicate that selecting for increased genistein content may also indirectly increase yield. The reverse of this effect, yield selections with increased isoflavones, has been previously observed by Morrison et al. (2008).

Continued efforts should be made to increase genistein content due to the numerous associated benefits, as well as the positive correlation with yield and total isoflavone content. While the results from this study indicate that phenotypic selections for genistein outperformed MAS, more research should be done comparing these selection methods. Soybean breeding with MAS will continue to gain more prominence as high throughput marker technology becomes cheaper and more widespread. The costly and time consuming process of collecting phenotypic data for isoflavones provides incentive to pursue MAS as an improvement strategy.
References
References


Soybeans: Improvement, Production, and Uses 3rd Edition. ASA, CSSA, and SSSA, Madison, WI.


CHAPTER 4
CONCLUSION AND FUTURE RESEARCH
Conclusion and Future Research

This study sought to detect and validate QTL for soybean \( \text{Glycine max (L.) Merrill} \) isoflavones genistein, daidzein, glycine, and total isoflavone content, and to use the genistein QTL for MAS to be compared with phenotypic selections. Overall, 21 QTL were detected for soybean isoflavones, including 7 for genistein, 5 for daidzein, 3 for glycine, and 6 for total isoflavones. Of these 21 QTL, 8 were newly detected, while 13 were validated from previous studies. The newly detected QTL will help to provide an improved understanding of the genetic regions controlling isoflavones. Of particular interest among the newly detected QTL were those whose \( R^2 \) values were at least 0.10, including GEN2, GEN4, GEN5 (Table 2.3), DAI5 (Table 2.4), and ISO2 (Table 2.6). These findings represent newly detected major QTL for soybean isoflavones. The 13 validated QTL also represent important findings, as they were originally detected in different environments than those detected in this study. These validated QTL had been previously detected in studies using both different parents and with ‘Essex’ as a common parent. The validation of these QTL in different as well as similar genetic populations represents a noteworthy consistency.

The QTL detected for genistein were used for MAS, which were compared with phenotypic selections for genistein. The low MAS selection from the Gen4eMAS test was not significantly different from the low phenotypic selection in the Gen4e test. Overall, the high phenotypic selections outperformed the MAS for genistein. However, the high MAS for genistein did show improvements over the high parent (‘Essex’), as well as improvements over unselected RILs. Additional results from this study indicated that genistein was positively correlated with other soybean isoflavones, as well as with yield. These findings are encouraging for the potential of MAS for soybean isoflavones, which is important as phenotyping for isoflavones is costly and time consuming.

Experience gained from the completion of this study provided insight into considerations for future research. This experiment suffered a nearly fatal blow when the original MAS, conducted in the spring of 2010, were discovered to have been done for height and not for genistein after two years of planting. Avoiding similar mistakes in future research will be emphasized with thorough evaluation of data.
While this research was useful in comparing MAS with phenotypic selections for genistein, future research would consider this comparison for all isoflavones. Additionally, a much more liberal threshold for making selections would be imposed. Using the example of a 20% phenotypic selection threshold, comparison with MAS could be made for the top 5%, the top 10%, and the top 20% of phenotypic selections. This would provide valuable information on the comparison of selection methods.

Additionally, efforts would be made to include all RILs which were candidates for MAS as entries in field tests. With the inclusion of all MAS candidates, comparisons could be made with phenotypic selections to determine which method predicted the overall top performer. Such information would provide insight into which method would be preferred in order to achieve maximum genetic gain.

Further considerations of future research will include emphasis on accurate and low cost techniques for phenotypic analysis of agronomic and seed quality traits. Also, continued research on molecular techniques for plant breeding will be considered. The increase in low cost high throughput molecular technology will continue to advance the potential of MAS for plant breeding.
APPENDIX A: TABLES
Table 2.1 Descriptive statistics and heritability values for a population of 274 RILs derived from ‘Essex’ and ‘Williams 82’, grown over three environments (Knoxville, TN; Harrisburg, IL; and Stuttgart, AR) with three replications. Parental LSMEANS also included.

<table>
<thead>
<tr>
<th>Test</th>
<th>Trait</th>
<th>Min⁹</th>
<th>Mean⁹</th>
<th>Max⁹</th>
<th>Dev. Essex ⁹</th>
<th>Williams ⁸² ⁹</th>
<th>h² ⁹</th>
</tr>
</thead>
<tbody>
<tr>
<td>Early</td>
<td>Genistein</td>
<td>0.44</td>
<td>0.69</td>
<td>0.96</td>
<td>0.12</td>
<td>0.77</td>
<td>0.74</td>
</tr>
<tr>
<td></td>
<td>Daidzein</td>
<td>0.62</td>
<td>0.80</td>
<td>0.96</td>
<td>0.08</td>
<td>0.86</td>
<td>0.84</td>
</tr>
<tr>
<td></td>
<td>Glycitein</td>
<td>0.15</td>
<td>0.17</td>
<td>0.19</td>
<td>0.01</td>
<td>0.16</td>
<td>0.18</td>
</tr>
<tr>
<td></td>
<td>Total Isoflavones</td>
<td>1.21</td>
<td>1.66</td>
<td>2.05</td>
<td>0.20</td>
<td>1.80</td>
<td>1.75</td>
</tr>
<tr>
<td>Mid</td>
<td>Genistein</td>
<td>0.51</td>
<td>0.75</td>
<td>0.93</td>
<td>0.09</td>
<td>0.84</td>
<td>0.62</td>
</tr>
<tr>
<td></td>
<td>Daidzein</td>
<td>0.69</td>
<td>0.83</td>
<td>0.93</td>
<td>0.06</td>
<td>0.84</td>
<td>0.79</td>
</tr>
<tr>
<td></td>
<td>Glycitein</td>
<td>0.15</td>
<td>0.16</td>
<td>0.18</td>
<td>0.01</td>
<td>0.16</td>
<td>0.17</td>
</tr>
<tr>
<td></td>
<td>Total Isoflavones</td>
<td>1.39</td>
<td>1.74</td>
<td>2.02</td>
<td>0.15</td>
<td>1.84</td>
<td>1.58</td>
</tr>
<tr>
<td>Late</td>
<td>Genistein</td>
<td>0.40</td>
<td>0.68</td>
<td>0.89</td>
<td>0.08</td>
<td>0.70</td>
<td>0.42</td>
</tr>
<tr>
<td></td>
<td>Daidzein</td>
<td>0.61</td>
<td>0.78</td>
<td>0.93</td>
<td>0.05</td>
<td>0.78</td>
<td>0.64</td>
</tr>
<tr>
<td></td>
<td>Glycitein</td>
<td>0.15</td>
<td>0.16</td>
<td>0.18</td>
<td>0.01</td>
<td>0.16</td>
<td>0.16</td>
</tr>
<tr>
<td></td>
<td>Total Isoflavones</td>
<td>1.16</td>
<td>1.62</td>
<td>1.98</td>
<td>0.13</td>
<td>1.64</td>
<td>1.22</td>
</tr>
</tbody>
</table>

⁹LSMEAN value expressed in mg g⁻¹ seed on a dry weight basis

⁹Heritability on an entry-mean basis
Table 2.2 Phenotypic correlations between genistein, daidzein, glycitein, and total isoflavones with agronomic and seed quality traits of interest in soybean from a population of 274 RILs derived from ‘Essex’ and ‘Williams 82’, grown over three environments (Knoxville, TN; Harrisburg, IL; and Stuttgart, AR) with three replications.

|                | Daidzein | Glycitein | Total  | Maturity | Lodging | Height | Yield | Protein | Oil   | Oleic Acid | Linoleic Acid | Linolenic Acid | Palmitic Acid | Stearic Acid | Oleic Acid | Linoleic Acid | Linolenic Acid | P value   |
|----------------|----------|-----------|--------|----------|---------|--------|-------|---------|-------|------------|---------------|----------------|---------------|--------------|------------|-------------|---------------|-----------|----------|
| Genistein      | 0.95     | 0.11      | 0.99   | 0.38     | -0.17   | -0.23  | 0.39  | -0.74   | -0.38 | 0.34       | -0.42          | -0.53          | 0.50          | 0.51        | <.0001     | <.0001      | <.0001       | <.0001    |<.0001    |
| P value        | <.0001   | <.0001    | <.0001 | 0.005    | <.0001  | <.0001 | <.0001| <.0001  | <.0001| <.0001    | <.0001        | <.0001         | <.0001        | <.0001       | <.0001     | <.0001      | <.0001       | <.0001    |<.0001    |
| Daidzein       | 0.21     | 0.98      | 0.24   | -0.25    | -0.34   | 0.28   | -0.64 | -0.38   | 0.40  | -0.37      | -0.59          | 0.56           | 0.50          | 0.50        | <.0001     | <.0001      | <.0001       | <.0001    |<.0001    |
| P value        | 0.0005   | <.0001    | <.0001 | <.0001   | <.0001  | <.0001 | <.0001| <.0001  | <.0001| <.0001    | <.0001        | <.0001         | <.0001        | <.0001       | <.0001     | <.0001      | <.0001       | <.0001    |<.0001    |
| Glycitein      | 0.19     | -0.50     | -0.21  | -0.22    | -0.28   | 0.04   | 0.30  | 0.17    | 0.31  | 0.12       | -0.16          | -0.18         |             |             |            |             | <.0001       | <.0001    |<.0001    |
| P value        | 0.0013   | <.0001    | 0.0005 | 0.0002   | 0.4967  | 0.0049 | <.0001| 0.0477  | 0.0071| 0.003     |               |               |              |             |             |             | <.0001      | <.0001    |<.0001    |
| Total          | 0.30     | -0.21     | -0.28  | 0.34     | -0.70   | -0.37  | 0.37  | -0.39   | -0.55 | 0.52       | 0.50           |               |              |             |             |             | <.0001      | <.0001    |<.0001    |
| P value        | <.0001   | 0.0004    | <.0001 | <.0001   | <.0001  | <.0001 | <.0001| <.0001  | <.0001| <.0001    | <.0001        |               |              |             |             |             | <.0001      | <.0001    |<.0001    |
Table 2.3 Quantitative trait loci associated with the isoflavone genistein in individual maturity tests in an ‘Essex’ × ‘Williams 82’ Population of 274 RILs grown over three environments (Knoxville, TN; Harrisburg, IL; and Stuttgart, AR). GEN1-GEN7 represent 7 QTL for the isoflavone genistein.

<table>
<thead>
<tr>
<th>Test</th>
<th>Chr</th>
<th>LG</th>
<th>QTL</th>
<th>Position (cM)</th>
<th>LOD score</th>
<th>Closest Marker</th>
<th>LOD - 1 Interval</th>
<th>R^2</th>
<th>Effect^b</th>
</tr>
</thead>
<tbody>
<tr>
<td>Early</td>
<td>6</td>
<td>C2</td>
<td>GEN1c</td>
<td>209.4</td>
<td>7.9</td>
<td>BARC-023517-05442</td>
<td>208.0-214.0</td>
<td>0.16</td>
<td>0.055</td>
</tr>
<tr>
<td></td>
<td>9</td>
<td>K</td>
<td>GEN2</td>
<td>95.1</td>
<td>10.5</td>
<td>BARC-038909-07393</td>
<td>86.1-103.0</td>
<td>0.23</td>
<td>0.058</td>
</tr>
<tr>
<td></td>
<td>13b</td>
<td>F</td>
<td>GEN3</td>
<td>32.4</td>
<td>6.0</td>
<td>BARC-030359-06859</td>
<td>26.6-40.3</td>
<td>0.14</td>
<td>-0.046</td>
</tr>
<tr>
<td>Mid</td>
<td>5</td>
<td>A1</td>
<td>GEN4</td>
<td>186.3</td>
<td>4.0</td>
<td>BARC-042853-08438</td>
<td>182.8-189.9</td>
<td>0.13</td>
<td>0.034</td>
</tr>
<tr>
<td></td>
<td>17</td>
<td>D2</td>
<td>GEN5</td>
<td>21.6</td>
<td>3.2</td>
<td>BARC-030909-06973</td>
<td>21.0-34.2</td>
<td>0.10</td>
<td>-0.029</td>
</tr>
<tr>
<td></td>
<td>19</td>
<td>L</td>
<td>GEN6d</td>
<td>137.2</td>
<td>7.6</td>
<td>BARC-035235-07156</td>
<td>132.1-148.8</td>
<td>0.22</td>
<td>0.044</td>
</tr>
<tr>
<td>Late</td>
<td>6</td>
<td>C2</td>
<td>GEN1c</td>
<td>205.9</td>
<td>3.1</td>
<td>BARC-066175-19800</td>
<td>203.9-212.9</td>
<td>0.09</td>
<td>0.061</td>
</tr>
<tr>
<td></td>
<td>9</td>
<td>K</td>
<td>GEN7</td>
<td>48.3</td>
<td>3.3</td>
<td>BARC-048623-10678</td>
<td>48.1-52.1</td>
<td>0.09</td>
<td>0.027</td>
</tr>
<tr>
<td></td>
<td>19</td>
<td>L</td>
<td>GEN6d</td>
<td>138.8</td>
<td>3.3</td>
<td>BARC-016145-02292</td>
<td>137.3-140.9</td>
<td>0.09</td>
<td>0.029</td>
</tr>
</tbody>
</table>

^aCm positions are those mapped in this population.
^bThe additive effect with respect to the 'Essex' allele expressed in mg genistein g^-1 of seed on a dry weight basis.
^cClosely associated with E1 locus.
^dClosely associated with Dt1 locus.
Table 2.4 Quantitative trait loci associated with the isoflavone daidzein in individual maturity tests in an ‘Essex’ × ‘Williams 82’ Population of 274 RILs grown over three environments (Knoxville, TN; Harrisburg, IL; and Stuttgart, AR). DAI1-DAI5 represent 5 QTL for the isoflavone daidzein.

<table>
<thead>
<tr>
<th>Test</th>
<th>Chr</th>
<th>LG</th>
<th>QTL</th>
<th>Position (cM)</th>
<th>LOD score</th>
<th>Closest Marker</th>
<th>LOD - 1 Interval</th>
<th>R²</th>
<th>Effect</th>
</tr>
</thead>
<tbody>
<tr>
<td>Early</td>
<td>6</td>
<td>C2</td>
<td>DAI1</td>
<td>209.4</td>
<td>5.8</td>
<td>BARC-023517-05442</td>
<td>208.0-214.1</td>
<td>0.13</td>
<td>0.033</td>
</tr>
<tr>
<td></td>
<td>9</td>
<td>K</td>
<td>DAI2</td>
<td>93.4</td>
<td>7.1</td>
<td>BARC-038909-07393</td>
<td>84.4-99.6</td>
<td>0.16</td>
<td>0.032</td>
</tr>
<tr>
<td></td>
<td>13b</td>
<td>F</td>
<td>DAI3</td>
<td>28.7</td>
<td>4.3</td>
<td>BARC-030359-06859</td>
<td>24.1-38.8</td>
<td>0.11</td>
<td>-0.026</td>
</tr>
<tr>
<td></td>
<td>19</td>
<td>L</td>
<td>DAI4</td>
<td>141.1</td>
<td>4.8</td>
<td>BARC-024345-04854</td>
<td>126.5-141.9</td>
<td>0.10</td>
<td>0.029</td>
</tr>
<tr>
<td>Mid</td>
<td>5</td>
<td>A1</td>
<td>DAI5</td>
<td>184.2</td>
<td>4.8</td>
<td>BARC-042853-08438</td>
<td>181.6-188.5</td>
<td>0.16</td>
<td>0.023</td>
</tr>
<tr>
<td></td>
<td>19</td>
<td>L</td>
<td>DAI4</td>
<td>144.5</td>
<td>8.0</td>
<td>BARC-026069-05243</td>
<td>135.6-149.5</td>
<td>0.30</td>
<td>0.032</td>
</tr>
<tr>
<td>Late</td>
<td>9</td>
<td>K</td>
<td>DAI2</td>
<td>98.0</td>
<td>4.1</td>
<td>BARC-038909-07393</td>
<td>82.9-120.5</td>
<td>0.13</td>
<td>0.019</td>
</tr>
<tr>
<td></td>
<td>19</td>
<td>L</td>
<td>DAI4</td>
<td>138.8</td>
<td>6.3</td>
<td>BARC-016145-02292</td>
<td>129.6-148.6</td>
<td>0.18</td>
<td>0.024</td>
</tr>
</tbody>
</table>

a. cM positions are those mapped in this population.
b. The additive effect with respect to the ‘Essex’ allele expressed in mg daidzein g⁻¹ of seed on a dry weight basis.
c. Closely associated with E1 locus.
d. Closely associated with Dt1 locus.
Table 2.5 Quantitative trait loci associated with the isoflavone glycitein in individual maturity tests in an ‘Essex’ × ‘Williams 82’ Population of 274 RILs grown over three environments (Knoxville, TN; Harrisburg, IL; and Stuttgart, AR). GLY1-GLY3 represent 3 QTL for the isoflavone glycine.

<table>
<thead>
<tr>
<th>Test</th>
<th>Chr</th>
<th>LG</th>
<th>QTL</th>
<th>Position (cM)</th>
<th>LOD score</th>
<th>Closest Marker</th>
<th>LOD - 1 Interval</th>
<th>R²</th>
<th>Effect</th>
</tr>
</thead>
<tbody>
<tr>
<td>Early</td>
<td>6</td>
<td>C2</td>
<td>GLY1</td>
<td>200.0</td>
<td>4.2</td>
<td>BARC-031337-07051</td>
<td>187.9-206.4</td>
<td>0.15</td>
<td>-0.003</td>
</tr>
<tr>
<td>Mid</td>
<td>9</td>
<td>K</td>
<td>GLY2</td>
<td>77.6</td>
<td>4.6</td>
<td>BARC-014813-01678</td>
<td>65.3-94.5</td>
<td>0.20</td>
<td>0.002</td>
</tr>
<tr>
<td></td>
<td>20</td>
<td>I</td>
<td>GLY3</td>
<td>105.0</td>
<td>3.5</td>
<td>BARC-053725-11957</td>
<td>90.0-113.5</td>
<td>0.17</td>
<td>-0.002</td>
</tr>
<tr>
<td>Late</td>
<td>NONE</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

a cM positions are those mapped in this population.

b The additive effect with respect to the ‘Essex’ allele expressed in mg glycitein g⁻¹ of seed on a dry weight basis.

c Closely associated with E1 locus.
Table 2.6 Quantitative trait loci associated with total isoflavones in individual maturity tests in an ‘Essex’ × ‘Williams 82’ Population of 274 RILs grown over three environments (Knoxville, TN; Harrisburg, IL; and Stuttgart, AR). ISO1-ISO6 represent 6 QTL for total isoflavones.

<table>
<thead>
<tr>
<th>Test</th>
<th>Chr</th>
<th>LG</th>
<th>QTL</th>
<th>Position (cM)</th>
<th>LOD score</th>
<th>Closest Marker</th>
<th>LOD - 1 Interval</th>
<th>R²</th>
<th>Effect</th>
</tr>
</thead>
<tbody>
<tr>
<td>Early</td>
<td>6</td>
<td>C2</td>
<td>ISO1</td>
<td>209.4</td>
<td>6.4</td>
<td>BARC-023517-05442</td>
<td>208.2-213.6</td>
<td>0.13</td>
<td>0.085</td>
</tr>
<tr>
<td></td>
<td>9</td>
<td>K</td>
<td>ISO2</td>
<td>94.7</td>
<td>9.3</td>
<td>BARC-038909-07393</td>
<td>86.3-102.5</td>
<td>0.21</td>
<td>0.092</td>
</tr>
<tr>
<td></td>
<td>13b</td>
<td>F</td>
<td>ISO3</td>
<td>31.4</td>
<td>4.8</td>
<td>BARC-030359-06859</td>
<td>25.0-41.2</td>
<td>0.11</td>
<td>-0.068</td>
</tr>
<tr>
<td></td>
<td>19</td>
<td>L</td>
<td>ISO4</td>
<td>141.1</td>
<td>4.5</td>
<td>BARC-024345-04854</td>
<td>123.7-142.0</td>
<td>0.09</td>
<td>0.068</td>
</tr>
<tr>
<td>Mid</td>
<td>5</td>
<td>A1</td>
<td>ISO5</td>
<td>187.2</td>
<td>4.3</td>
<td>BARC-059081-15595</td>
<td>182.6-191.7</td>
<td>0.09</td>
<td>0.045</td>
</tr>
<tr>
<td></td>
<td>19</td>
<td>L</td>
<td>ISO4</td>
<td>137.3</td>
<td>6.1</td>
<td>BARC-035235-07156</td>
<td>130.5-148.5</td>
<td>0.18</td>
<td>0.062</td>
</tr>
<tr>
<td>Late</td>
<td>5</td>
<td>A1</td>
<td>ISO6</td>
<td>26.6</td>
<td>3.3</td>
<td>BARC-019415-03923</td>
<td>14.7-50.6</td>
<td>0.09</td>
<td>0.041</td>
</tr>
<tr>
<td></td>
<td>19</td>
<td>L</td>
<td>ISO4</td>
<td>138.8</td>
<td>3.9</td>
<td>BARC-016145-02292</td>
<td>130.1-148.4</td>
<td>0.11</td>
<td>0.049</td>
</tr>
</tbody>
</table>

ᵃCm positions are those mapped in this population.
ᵇThe additive effect with respect to the ‘Essex’ allele expressed in mg total isoflavones g⁻¹ of seed on a dry weight basis.
ᶜClosely associated with E1 locus.
ᵈClosely associated with Dt1 locus.
Table 3.1 Quantitative trait loci associated with height in individual maturity tests in an ‘Essex’ × ‘Williams 82’ Population of 274 RILs grown over three environments (Knoxville, TN; Harrisburg, IL; and Stuttgart, AR). HGT1 represents 1 QTL for soybean height.

<table>
<thead>
<tr>
<th>Test</th>
<th>Chr</th>
<th>LG</th>
<th>QTL</th>
<th>Position (cM)</th>
<th>LOD score</th>
<th>Closest Marker</th>
<th>LOD - 1 Interval</th>
<th>( R^2 )</th>
<th>Effect(^b)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Early</td>
<td>19</td>
<td>L</td>
<td>HGT1(^c)</td>
<td>139.0</td>
<td>9.9</td>
<td>BARC-016145-02292</td>
<td>137.9-140.8</td>
<td>0.35</td>
<td>-6.23</td>
</tr>
<tr>
<td>Mid</td>
<td>19</td>
<td>L</td>
<td>HGT1(^c)</td>
<td>138.2</td>
<td>32</td>
<td>BARC-016145-02292</td>
<td>137.9-138.8</td>
<td>0.77</td>
<td>-15.50</td>
</tr>
<tr>
<td>Late</td>
<td>19</td>
<td>L</td>
<td>HGT1(^c)</td>
<td>138.8</td>
<td>29.2</td>
<td>BARC-016145-02292</td>
<td>137.6-144.4</td>
<td>0.69</td>
<td>-15.07</td>
</tr>
</tbody>
</table>

\(^a\)Cm positions are those mapped in this population.
\(^b\)The additive effect with respect to the ‘Essex’ allele expressed in cm.
\(^c\)Closely associated with Dt1 locus.
Table 3.2 Comparison of height LSMEANS for lines in the 4eMAS, 4LMAS, and 5eMAS field tests from 2010 and 2011.

<table>
<thead>
<tr>
<th>Test</th>
<th>Line</th>
<th>Tail&lt;sup&gt;a&lt;/sup&gt;</th>
<th>Height (cm)</th>
<th>Letter Group&lt;sup&gt;b&lt;/sup&gt;</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gen4eMAS</td>
<td>ExW3-164</td>
<td>high</td>
<td>108.94</td>
<td>A</td>
</tr>
<tr>
<td></td>
<td>ExW3-203</td>
<td>high</td>
<td>103.01</td>
<td>A</td>
</tr>
<tr>
<td>Gen4LMAS</td>
<td>ExW3-239</td>
<td>high</td>
<td>116.30</td>
<td>A</td>
</tr>
<tr>
<td></td>
<td>ExW3-331</td>
<td>high</td>
<td>112.32</td>
<td>AB</td>
</tr>
<tr>
<td></td>
<td>ExW3-144</td>
<td>high</td>
<td>110.09</td>
<td>AB</td>
</tr>
<tr>
<td></td>
<td>ExW3-390</td>
<td>high</td>
<td>105.01</td>
<td>AB</td>
</tr>
<tr>
<td></td>
<td>ExW3-614</td>
<td>high</td>
<td>105.01</td>
<td>AB</td>
</tr>
<tr>
<td></td>
<td>ExW3-332</td>
<td>high</td>
<td>104.73</td>
<td>AB</td>
</tr>
<tr>
<td></td>
<td>ExW3-222</td>
<td>high</td>
<td>101.91</td>
<td>B</td>
</tr>
<tr>
<td></td>
<td>ExW3-316</td>
<td>low</td>
<td>83.11</td>
<td>C</td>
</tr>
<tr>
<td></td>
<td>ExW3-069</td>
<td>low</td>
<td>82.97</td>
<td>C</td>
</tr>
<tr>
<td></td>
<td>ExW3-651</td>
<td>low</td>
<td>80.74</td>
<td>CD</td>
</tr>
<tr>
<td></td>
<td>ExW3-020</td>
<td>low</td>
<td>75.64</td>
<td>CDE</td>
</tr>
<tr>
<td></td>
<td>ExW3-685</td>
<td>low</td>
<td>69.45</td>
<td>DEF</td>
</tr>
<tr>
<td></td>
<td>ExW3-552</td>
<td>low</td>
<td>67.73</td>
<td>EF</td>
</tr>
<tr>
<td></td>
<td>ExW3-675</td>
<td>low</td>
<td>63.64</td>
<td>F</td>
</tr>
<tr>
<td>Gen5eMAS</td>
<td>ExW3-467</td>
<td>high</td>
<td>129.68</td>
<td>A</td>
</tr>
<tr>
<td></td>
<td>ExW3-411</td>
<td>high</td>
<td>125.65</td>
<td>AB</td>
</tr>
<tr>
<td></td>
<td>ExW3-446</td>
<td>high</td>
<td>117.46</td>
<td>ABC</td>
</tr>
<tr>
<td></td>
<td>ExW3-437</td>
<td>high</td>
<td>116.05</td>
<td>ABC</td>
</tr>
<tr>
<td></td>
<td>ExW3-329</td>
<td>high</td>
<td>113.79</td>
<td>ABC</td>
</tr>
<tr>
<td></td>
<td>ExW3-447</td>
<td>high</td>
<td>109.56</td>
<td>BC</td>
</tr>
<tr>
<td></td>
<td>ExW3-079</td>
<td>high</td>
<td>108.43</td>
<td>BC</td>
</tr>
<tr>
<td></td>
<td>ExW3-177</td>
<td>high</td>
<td>106.74</td>
<td>BC</td>
</tr>
<tr>
<td></td>
<td>ExW3-473</td>
<td>high</td>
<td>103.86</td>
<td>C</td>
</tr>
<tr>
<td></td>
<td>ExW3-324</td>
<td>low</td>
<td>78.18</td>
<td>D</td>
</tr>
<tr>
<td></td>
<td>ExW3-096</td>
<td>low</td>
<td>78.05</td>
<td>D</td>
</tr>
<tr>
<td></td>
<td>‘Essex’ parent</td>
<td></td>
<td>75.78</td>
<td>D</td>
</tr>
</tbody>
</table>

<sup>a</sup>Height marker-assisted selection

<sup>b</sup>Tukey-Kramer mean separation; LSMEANS followed by the same letter within a test are not significantly different at the 5% level of probability
Table 3.3 Comparison of height LSMEANS for tail selections in the 4LMAS and 5eMAS field tests from 2010 and 2011.

<table>
<thead>
<tr>
<th>Test</th>
<th>Tail</th>
<th>Height (cm)</th>
<th>Letter Group</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gen4LMAS</td>
<td>high</td>
<td>107.91</td>
<td>A</td>
</tr>
<tr>
<td></td>
<td>low</td>
<td>74.76</td>
<td>B</td>
</tr>
<tr>
<td>Gen5eMAS</td>
<td>high</td>
<td>114.58</td>
<td>A</td>
</tr>
<tr>
<td></td>
<td>low</td>
<td>78.11</td>
<td>B</td>
</tr>
<tr>
<td></td>
<td>Essex</td>
<td>75.78</td>
<td>B</td>
</tr>
</tbody>
</table>

\(^a\)Height marker-assisted selection

\(^b\)Tukey-Kramer mean separation; LSMEANS followed by the same letter within a test are not significantly different at the 5% level of probability
Table 3.4 Comparison of genistein LSMEANS for lines in the 3L, 4e, and 5 Late maturity groups from 2010 and 2011.

<table>
<thead>
<tr>
<th>Maturity Group</th>
<th>Line</th>
<th>Test</th>
<th>Tail&lt;sup&gt;a&lt;/sup&gt;</th>
<th>Genistein&lt;sup&gt;b&lt;/sup&gt;</th>
<th>Letter</th>
<th>Group&lt;sup&gt;b&lt;/sup&gt;</th>
</tr>
</thead>
<tbody>
<tr>
<td>3 Late</td>
<td>ExW3-519</td>
<td>GEN 3L</td>
<td>high</td>
<td>0.53</td>
<td>A</td>
<td></td>
</tr>
<tr>
<td></td>
<td>‘Williams 82’</td>
<td>GEN 3L</td>
<td>parent</td>
<td>0.39</td>
<td>B</td>
<td></td>
</tr>
<tr>
<td></td>
<td>ExW3-612</td>
<td>GEN 3L</td>
<td>low</td>
<td>0.28</td>
<td>C</td>
<td></td>
</tr>
<tr>
<td>4 Early</td>
<td>ExW3-203</td>
<td>GEN4eMAS</td>
<td>unselected</td>
<td>0.61</td>
<td>A</td>
<td></td>
</tr>
<tr>
<td></td>
<td>ExW3-635&lt;sup&gt;c&lt;/sup&gt;</td>
<td>GEN4e</td>
<td>high</td>
<td>0.55</td>
<td>A</td>
<td></td>
</tr>
<tr>
<td></td>
<td>ExW3-618&lt;sup&gt;d&lt;/sup&gt;</td>
<td>GEN4e</td>
<td>low</td>
<td>0.38</td>
<td>B</td>
<td></td>
</tr>
<tr>
<td></td>
<td>ExW3-164</td>
<td>GEN4eMAS</td>
<td>low</td>
<td>0.36</td>
<td>B</td>
<td></td>
</tr>
<tr>
<td>4 Late</td>
<td>ExW3-457&lt;sup&gt;c&lt;/sup&gt;</td>
<td>GEN4L</td>
<td>high</td>
<td>0.74</td>
<td>A</td>
<td></td>
</tr>
<tr>
<td></td>
<td>ExW3-073</td>
<td>GEN4L</td>
<td>high</td>
<td>0.70</td>
<td>AB</td>
<td></td>
</tr>
<tr>
<td></td>
<td>ExW3-069</td>
<td>GEN4LMAS</td>
<td>high</td>
<td>0.68</td>
<td>AB</td>
<td></td>
</tr>
<tr>
<td></td>
<td>ExW3-685</td>
<td>GEN4LMAS</td>
<td>high</td>
<td>0.67</td>
<td>ABC</td>
<td></td>
</tr>
<tr>
<td></td>
<td>ExW3-239</td>
<td>GEN4LMAS</td>
<td>unselected</td>
<td>0.66</td>
<td>ABC</td>
<td></td>
</tr>
<tr>
<td></td>
<td>ExW3-651</td>
<td>GEN4LMAS</td>
<td>unselected</td>
<td>0.65</td>
<td>ABC</td>
<td></td>
</tr>
<tr>
<td></td>
<td>ExW3-331</td>
<td>GEN4LMAS</td>
<td>unselected</td>
<td>0.64</td>
<td>BC</td>
<td></td>
</tr>
<tr>
<td></td>
<td>ExW3-675</td>
<td>GEN4LMAS</td>
<td>high</td>
<td>0.61</td>
<td>BC</td>
<td></td>
</tr>
<tr>
<td></td>
<td>ExW3-385</td>
<td>GEN4L</td>
<td>high</td>
<td>0.61</td>
<td>BC</td>
<td></td>
</tr>
<tr>
<td></td>
<td>ExW3-614</td>
<td>GEN4LMAS</td>
<td>unselected</td>
<td>0.60</td>
<td>BCD</td>
<td></td>
</tr>
<tr>
<td></td>
<td>ExW3-332</td>
<td>GEN4LMAS</td>
<td>unselected</td>
<td>0.60</td>
<td>BCD</td>
<td></td>
</tr>
<tr>
<td></td>
<td>ExW3-390</td>
<td>GEN4LMAS</td>
<td>unselected</td>
<td>0.60</td>
<td>BCD</td>
<td></td>
</tr>
<tr>
<td></td>
<td>ExW3-614</td>
<td>GEN4L</td>
<td>low</td>
<td>0.58</td>
<td>CD</td>
<td></td>
</tr>
<tr>
<td></td>
<td>ExW3-552</td>
<td>GEN4LMAS</td>
<td>unselected</td>
<td>0.57</td>
<td>CD</td>
<td></td>
</tr>
<tr>
<td></td>
<td>ExW3-020</td>
<td>GEN4LMAS</td>
<td>high</td>
<td>0.56</td>
<td>CD</td>
<td></td>
</tr>
<tr>
<td></td>
<td>ExW3-144</td>
<td>GEN4LMAS</td>
<td>unselected</td>
<td>0.55</td>
<td>CD</td>
<td></td>
</tr>
<tr>
<td></td>
<td>ExW3-222</td>
<td>GEN4LMAS</td>
<td>unselected</td>
<td>0.55</td>
<td>CD</td>
<td></td>
</tr>
<tr>
<td></td>
<td>ExW3-065</td>
<td>GEN4L</td>
<td>low</td>
<td>0.49</td>
<td>D</td>
<td></td>
</tr>
<tr>
<td></td>
<td>ExW3-316</td>
<td>GEN4LMAS</td>
<td>unselected</td>
<td>0.49</td>
<td>D</td>
<td></td>
</tr>
<tr>
<td></td>
<td>ExW3-258</td>
<td>GEN4L</td>
<td>low</td>
<td>0.39</td>
<td>E</td>
<td></td>
</tr>
</tbody>
</table>

<sup>a</sup>Genistein marker-assisted selection  
<sup>b</sup>Tukey-Kramer mean separation; LSMEANS followed by the same letter within a test are not significantly different at the 5% level of probability  
<sup>c</sup>Also a candidate for high marker-selection  
<sup>d</sup>Also a candidate for low marker-selection
Table 3.5 Comparison of genistein LSMEANS for tail selections in the 3L, 4e, 4L, and 5e Maturity Groups from 2010 and 2011.

<table>
<thead>
<tr>
<th>Maturity Group</th>
<th>Test</th>
<th>Tail(^a)</th>
<th>Genistein (mg g(^{-1}))</th>
<th>Letter Group(^b)</th>
</tr>
</thead>
<tbody>
<tr>
<td>3 Late</td>
<td>GEN 3L</td>
<td>high</td>
<td>0.53</td>
<td>A</td>
</tr>
<tr>
<td></td>
<td>GEN 3L</td>
<td>'Williams 82'</td>
<td>0.39</td>
<td>B</td>
</tr>
<tr>
<td></td>
<td>GEN 3L</td>
<td>low</td>
<td>0.28</td>
<td>C</td>
</tr>
<tr>
<td>4 Early</td>
<td>GEN4eMAS</td>
<td>unselected</td>
<td>0.61</td>
<td>A</td>
</tr>
<tr>
<td></td>
<td>GEN4e</td>
<td>high</td>
<td>0.55</td>
<td>A</td>
</tr>
<tr>
<td></td>
<td>GEN4e</td>
<td>low</td>
<td>0.38</td>
<td>B</td>
</tr>
<tr>
<td></td>
<td>GEN4eMAS</td>
<td>low</td>
<td>0.36</td>
<td>B</td>
</tr>
<tr>
<td>4 Late</td>
<td>GEN4L</td>
<td>high</td>
<td>0.68</td>
<td>A</td>
</tr>
<tr>
<td></td>
<td>GEN4LMAS</td>
<td>high</td>
<td>0.63</td>
<td>B</td>
</tr>
<tr>
<td></td>
<td>GEN4LMAS</td>
<td>unselected</td>
<td>0.59</td>
<td>C</td>
</tr>
<tr>
<td></td>
<td>GEN4L</td>
<td>low</td>
<td>0.49</td>
<td>D</td>
</tr>
<tr>
<td>5 Early</td>
<td>GEN5e</td>
<td>high</td>
<td>0.73</td>
<td>A</td>
</tr>
<tr>
<td></td>
<td>GEN5eMAS</td>
<td>high</td>
<td>0.62</td>
<td>B</td>
</tr>
<tr>
<td></td>
<td>GEN5eMAS</td>
<td>'Essex'</td>
<td>0.61</td>
<td>B</td>
</tr>
<tr>
<td></td>
<td>GEN5e</td>
<td>'Essex'</td>
<td>0.60</td>
<td>B</td>
</tr>
<tr>
<td></td>
<td>GEN5eMAS</td>
<td>unselected</td>
<td>0.52</td>
<td>C</td>
</tr>
<tr>
<td></td>
<td>GEN5e</td>
<td>low</td>
<td>0.45</td>
<td>D</td>
</tr>
</tbody>
</table>

\(^a\) Genistein marker-assisted selection

\(^b\) Tukey-Kramer mean separation; LSMEANS followed by the same letter within a test are not significantly different at the 5% level of probability
Table 3.6 Comparison of genistein for Lines in the 5e Maturity Group from 2010 and 2011.

<table>
<thead>
<tr>
<th>Maturity Group</th>
<th>Line</th>
<th>Test</th>
<th>Tail&lt;sup&gt;a&lt;/sup&gt;</th>
<th>Genistein (mg g&lt;sup&gt;-1&lt;/sup&gt;)</th>
<th>Letter</th>
<th>Group&lt;sup&gt;b&lt;/sup&gt;</th>
</tr>
</thead>
<tbody>
<tr>
<td>5 Early</td>
<td>ExW3-329</td>
<td>GEN5e</td>
<td>high</td>
<td>0.73</td>
<td>A</td>
<td></td>
</tr>
<tr>
<td></td>
<td>ExW3-231</td>
<td>GEN5e</td>
<td>high</td>
<td>0.73</td>
<td>A</td>
<td></td>
</tr>
<tr>
<td></td>
<td>ExW3-329</td>
<td>GEN5eMAS</td>
<td>unselected</td>
<td>0.71</td>
<td>AB</td>
<td></td>
</tr>
<tr>
<td></td>
<td>ExW3-096</td>
<td>GEN5eMAS</td>
<td>high</td>
<td>0.63</td>
<td>BC</td>
<td></td>
</tr>
<tr>
<td></td>
<td>ExW3-324</td>
<td>GEN5eMAS</td>
<td>high</td>
<td>0.61</td>
<td>BC</td>
<td></td>
</tr>
<tr>
<td></td>
<td>'Essex'</td>
<td>GEN5eMAS</td>
<td>parent</td>
<td>0.61</td>
<td>BC</td>
<td></td>
</tr>
<tr>
<td></td>
<td>ExW3-447</td>
<td>GEN5eMAS</td>
<td>unselected</td>
<td>0.60</td>
<td>BCD</td>
<td></td>
</tr>
<tr>
<td></td>
<td>'Essex'</td>
<td>GEN5e</td>
<td>parent</td>
<td>0.60</td>
<td>C</td>
<td></td>
</tr>
<tr>
<td></td>
<td>ExW3-446</td>
<td>GEN5eMAS</td>
<td>unselected</td>
<td>0.55</td>
<td>CDE</td>
<td></td>
</tr>
<tr>
<td></td>
<td>ExW3-079</td>
<td>GEN5eMAS</td>
<td>unselected</td>
<td>0.54</td>
<td>CDE</td>
<td></td>
</tr>
<tr>
<td></td>
<td>ExW3-177</td>
<td>GEN5eMAS</td>
<td>unselected</td>
<td>0.54</td>
<td>CDEF</td>
<td></td>
</tr>
<tr>
<td></td>
<td>ExW3-473</td>
<td>GEN5eMAS</td>
<td>unselected</td>
<td>0.50</td>
<td>DEF</td>
<td></td>
</tr>
<tr>
<td></td>
<td>ExW3-079</td>
<td>GEN5e</td>
<td>low</td>
<td>0.50</td>
<td>DEFG</td>
<td></td>
</tr>
<tr>
<td></td>
<td>ExW3-437</td>
<td>GEN5eMAS</td>
<td>unselected</td>
<td>0.46</td>
<td>EFGH</td>
<td></td>
</tr>
<tr>
<td></td>
<td>ExW3-467</td>
<td>GEN5eMAS</td>
<td>unselected</td>
<td>0.44</td>
<td>FGH</td>
<td></td>
</tr>
<tr>
<td></td>
<td>ExW3-411</td>
<td>GEN5e</td>
<td>low</td>
<td>0.41</td>
<td>GH</td>
<td></td>
</tr>
<tr>
<td></td>
<td>ExW3-411</td>
<td>GEN5eMAS</td>
<td>unselected</td>
<td>0.36</td>
<td>H</td>
<td></td>
</tr>
</tbody>
</table>

<sup>a</sup>Genistein marker-assisted selection  
<sup>b</sup>Tukey-Kramer mean separation; LSMEANS followed by the same letter within a test are not significantly different at the 5% level of probability
Table 3.7 Phenotypic correlations between genistein and agronomic and seed quality traits of interest in soybean from LSMEANS of 41 RILs derived from ‘Essex’ and ‘Williams 82’ grown in seven field tests (Gen3L, Gen4e, Gen4L, Gen5e, Gen4eMAS, Gen4LMAS, and Gen5eMAS). Field tests were grown in three environments (Knoxville, TN; Springfield, TN; and Milan, TN) with three replications over two years (2010 and 2011).

<table>
<thead>
<tr>
<th>Daidzein</th>
<th>Glycitein</th>
<th>Total</th>
<th>Maturity</th>
<th>Lodging</th>
<th>Height</th>
<th>Yield</th>
<th>Protein</th>
<th>Oil</th>
<th>Palmitic Acid</th>
<th>Stearic Acid</th>
<th>Oleic Acid</th>
<th>Linoleic Acid</th>
<th>Linolenic Acid</th>
</tr>
</thead>
<tbody>
<tr>
<td>Genistein</td>
<td>0.96</td>
<td>0.61</td>
<td>0.99</td>
<td>0.32</td>
<td>-0.20</td>
<td>-0.20</td>
<td>0.67</td>
<td>-0.83</td>
<td>-0.03</td>
<td>-0.54</td>
<td>-0.75</td>
<td>0.69</td>
<td>0.80</td>
</tr>
<tr>
<td>P value</td>
<td>&lt;.0001</td>
<td>&lt;.0001</td>
<td>&lt;.0001</td>
<td>0.0413</td>
<td>0.2102</td>
<td>0.1987</td>
<td>&lt;.0001</td>
<td>&lt;.0001</td>
<td>0.8645</td>
<td>0.0052</td>
<td>&lt;.0001</td>
<td>&lt;.0001</td>
<td>&lt;.0001</td>
</tr>
</tbody>
</table>
Figure 1.1 Chemical structures of soybean isoflavone aglycones (Pan et al., 2001)

<table>
<thead>
<tr>
<th>Compounds</th>
<th>R1</th>
<th>R2</th>
</tr>
</thead>
<tbody>
<tr>
<td>genistein</td>
<td>OH</td>
<td>H</td>
</tr>
<tr>
<td>daidzein</td>
<td>H</td>
<td>H</td>
</tr>
<tr>
<td>glycitein</td>
<td>H</td>
<td>OCH₃</td>
</tr>
</tbody>
</table>
Figure 1.2 Structural similarity between isoflavones and estrogen (Setchell and Cassidy, 1999).
Figure 2.1 Distributions of genistein and daidzein means from 274 F5 derived RILs of ‘Essex’ × ‘Williams 82’ separated into individual tests based on maturity, with each test containing 91 or 92 RILs. Field tests were grown in 2009 in Knoxville, TN; Harrisburg, IL; and Stuttgart, AR. All plots were harvested at maturity.
Figure 2.2 Distributions of glycine and total isoflavone means from 274 F5 derived RILs of ‘Essex’ × ‘Williams 82’ separated into individual tests based on maturity, with each test containing 91 or 92 RILs. Field tests were grown in 2009 in Knoxville, TN; Harrisburg, IL; and Stuttgart, AR. All plots were harvested at maturity.
Figure 2.3 Genetic map of ‘Essex’ × ‘Williams 82’ population of 274 RILs mapped with 480 polymorphic SNP markers. Chromosome 13 is split into 13a and 13b.
Figure 2.4 QTL positions for genistein (shown in red) from Early, Mid, and Late tests of 274 RILs of ‘Essex’ × ‘Williams 82’ grown over three environments (Knoxville, TN; Harrisburg, IL; and Stuttgart, AR). E1 maturity locus (chromosome 6, shown in green) and Dt1 growth habit locus (chromosome 19, shown in blue). The cM positions are from this population. The consensus map 4.0 positions of each BARC SNP are available at http://www.soybase.org/.
Figure 2.5 QTL positions for daidzein (shown in red) from Early, Mid, and Late tests of 274 RILs of ‘Essex’ × ‘Williams 82’ grown over three environments (Knoxville, TN; Harrisburg, IL; and Stuttgart, AR). E1 maturity locus (chromosome 6, shown in green) and Dt1 growth habit locus (chromosome 19, shown in blue). The cM positions are from this population. The consensus map 4.0 positions of each BARC SNP are available at http://www.soybase.org/.
Figure 2.6 QTL positions for glycitein (shown in red) from Early, Mid, and Late tests of 274 RILs of ‘Essex’ × ‘Williams 82’ grown over three environments (Knoxville, TN; Harrisburg, IL; and Stuttgart, AR). E1 maturity locus (chromosome 6, shown in green). The cM positions are from this population. The consensus map 4.0 positions of each BARC SNP are available at http://www.soybase.org/.
Figure 2.7 QTL positions for total isoflavones (shown in red) from Early, Mid, and Late tests of 274 RILs of ‘Essex’ × ‘Williams 82’ grown over three environments (Knoxville, TN; Harrisburg, IL; and Stuttgart, AR). E1 maturity locus (chromosome 6, shown in green) and Dt1 growth habit locus (chromosome 19, shown in blue). The cM positions are from this population. The consensus map 4.0 positions of each BARC SNP are available at http://www.soybase.org/.
Enzymes used in pathway: phenylalanine ammonia lyase (PAL), cinnamic acid 4-hydroxylase (C4H), 4-coumarate:CoA ligase (4CL), chalcone synthase (CHS), chalcone reductase (CHR), chalcone isomerase (CHI), isoflavone synthase (IFS), flavanone 3-hydroxylase (F3H), dihydroflavonol reductase (DFR) and isoflavone reductase (IFR).
Figure 3.1 Selected markers for height MAS for candidate RILs from the 2009 early test (92 F5 derived RILs) grown over three environments (Knoxville, TN; Harrisburg, IL; and Stuttgart, AR). LOD-1 interval used for selection (shown in red) and Dt1 growth habit locus (shown in blue) also displayed. The cM positions are from this population. The consensus map 4.0 positions of each BARC SNP are available at http://www.soybase.org/.
Figure 3.2 Selected markers for height MAS for candidate RILs from the 2009 mid test (91 F^5 derived RILs) grown over three environments (Knoxville, TN; Harrisburg, IL; and Stuttgart, AR). LOD-1 interval used for selection (shown in red) and Dt1 growth habit locus (shown in blue) also displayed. The cM positions are from this population. The consensus map 4.0 positions of each BARC SNP are available at http://www.soybase.org/.
Figure 3.3 Selected markers for height MAS for candidate RILs from the 2009 late test (91 F^5 derived RILs) grown over three environments (Knoxville, TN; Harrisburg, IL; and Stuttgart, AR). LOD-1 interval used for selection (shown in red) and Dt1 growth habit locus (shown in blue) also displayed. The cM positions are from this population. The consensus map 4.0 positions of each BARC SNP are available at http://www.soybase.org/.
Figure 3.4 Selected markers for genistein MAS for candidate RILs from the 2009 early test (92 F^5 derived RILs) grown over three environments (Knoxville, TN; Harrisburg, IL; and Stuttgart, AR). LOD-1 interval used for selection (shown in red) and E1 maturity locus (shown in green) also displayed. A QTL was also detected on chromosome 13b in the early test, but it was not used for selections as no markers were within the LOD-1 interval. The cM positions are from this population. The consensus map 4.0 positions of each BARC SNP are available at http://www.soybase.org/.
Figure 3.5 Selected markers for genistein MAS for candidate RILs from the 2009 mid test (91 F^5 derived RILs) grown over three environments (Knoxville, TN; Harrisburg, IL; and Stuttgart, AR). LOD-1 interval used for selection (shown in red) and Dt1 growth habit locus (shown in blue) also displayed. The cM positions are from this population. The consensus map 4.0 positions of each BARC SNP are available at http://www.soybase.org/.
Figure 3.6 Selected markers for genistein MAS for candidate RILs from the 2009 late test (91 F5 derived RILs) grown over three environments (Knoxville, TN; Harrisburg, IL; and Stuttgart, AR). LOD-1 interval used for selection (shown in red), E1 maturity locus (shown in green), and Dt1 growth habit locus (shown in blue) also displayed. The cM positions are from this population. The consensus map 4.0 positions of each BARC SNP are available at http://www.soybase.org/.
Figure 3.7 Comparison of LSMEAN heights of all tail selections in the 4LMAS and 5eMAS Field Tests from 2010 and 2011. Tukey-Kramer mean separation; LSMEANS followed by the same letter within a test are not significantly different at the 5% level of probability.
Figure 3.8 Comparison of genistein for LSMEAN tail selections in the 3L, 4e, and 4L Maturity Groups from 2010 and 2011. Tukey-Kramer mean separation; LSMEANS followed by the same letter within a test are not significantly different at the 5% level of probability. Tail selections from MAS tests shown in red and phenotypic tests shown in blue.
VITA

Christopher Joseph Smallwood was born in Laurel, MD, on the 11th of August, 1984. He attended the University of Tennessee, Knoxville, and received a Bachelor of Science in Wildlife & Fisheries Science, with a minor in Forestry in the spring of 2008.

In summer of 2010, he started a Master of Science program at the University of Tennessee, Knoxville in Plant Sciences, with a concentration in plant breeding. Upon completion of this degree, Chris hopes to continue his research in plant breeding at the University of Tennessee as a PhD student.