USE OF DIVERSE GERMPLASM TO IMPROVE PEANUT ROOT-KNOT NEMATODE RESISTANCE AND SEED PROTEIN CONTENT IN SOYBEAN

by

JENNIFER LYNN YATES

(Under the Direction of H. Roger Boerma)

ABSTRACT

Soybean is a major crop of which the USA is a leading world producer. The demand for soybean and soybean products continues to increase as its health benefits and industrial and food uses become more known; therefore, it is critical to improve soybean qualities of interest both to growers and processors to maintain this competitive edge in the international market. However, soybean cultivars currently grown in the USA have limited genetic diversity, and it is often necessary to explore unadapted germplasm to find novel genes for desirable traits. In this study, diverse germplasm sources from China and Korea were used as a source of genes for two desirable traits, peanut root-knot nematode (RKN) resistance and seed protein content. The resistance genes from three Chinese plant introductions (PIs) were evaluated for uniqueness of resistance genes to the peanut RKN, and the putative genes were identified through gene mapping. The gene identified in two of the three PIs was further found to be associated with peanut RKN resistance in southern soybean cultivars. The effect of a high-protein allele from a Korean cultivar on yield, protein and oil contents, and amino acid composition in three elite U.S. soybean cultivars was also assessed. Consistent effects of the high-protein gene on plant maturity, seed weight, protein and oil contents, and increases in the
concentrations of certain amino acids was discovered. However, the effect of the gene on yield was dependent on both the genetic background and the environment. This information aids soybean breeders in allocating their resources for managing emerging pests and in selecting soybean cultivars with improved protein quantity and quality profiles.

INDEX WORDS: soybean, germplasm, cultivar, peanut root-knot nematode, RKN, disease resistance, seed protein content, plant introduction, yield, oil, amino acid composition
USE OF DIVERSE GERMPLASM TO IMPROVE PEANUT ROOT-KNOT NEMATODE RESISTANCE AND SEED PROTEIN CONTENT IN SOYBEAN

by

JENNIFER LYNN YATES

B.S., University of Georgia, 1998
M.S., University of California-Davis, 2000

A Dissertation Submitted to the Graduate Faculty of The University of Georgia in Partial Fulfillment of the Requirements for the Degree

DOCTOR OF PHILOSOPHY

ATHENS, GEORGIA

2006
USE OF DIVERSE GERMLASM TO IMPROVE PEANUT ROOT-KNOT NEMATODE RESISTANCE AND SEED PROTEIN CONTENT IN SOYBEAN

by

JENNIFER LYNN YATES

Major Professor: H. Roger Boerma
Committee: Joseph Bouton
Katrien Devos
Richard Hussey
Wayne Parrott

Electronic Version Approved:
Maureen Grasso
Dean of the Graduate School
The University of Georgia
August 2006
DEDICATION

I especially dedicate this effort to my family, and particularly to Erika Swinson. Without your tireless help in running gel after gel, and without your patience, love, and good advice, I would not be where I am today.
ACKNOWLEDGEMENTS

This degree would not have been possible without the guidance and assistance of both my major professor and my colleagues in the lab. To Roger Boerma, for his patience, his mentorship, and his integrity, I feel lucky to have been your student. Special thanks goes to Dale Wood, Gina Rowan, Earl Baxter, Joe Head, and Herschel Chambers for help with field and greenhouse tests, and to Steve Finnerty and Charles Lester for assistance with the nematode greenhouse studies. Special thanks also to Jennie Alvernaz, David Hulburt, Erin Leggett, and Renée Hilton for all of their assistance in the lab.
# TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>ACKNOWLEDGEMENTS</td>
</tr>
<tr>
<td><strong>CHAPTER</strong></td>
</tr>
<tr>
<td>1</td>
</tr>
<tr>
<td>2</td>
</tr>
<tr>
<td>3</td>
</tr>
<tr>
<td>4</td>
</tr>
<tr>
<td>5</td>
</tr>
<tr>
<td>6</td>
</tr>
<tr>
<td>7</td>
</tr>
<tr>
<td>8</td>
</tr>
</tbody>
</table>
CHAPTER 1
INTRODUCTION

“The Soy-bean bears the climate of Pennsylvania very well. The bean ought therefore to be cultivated.”  J. Mease, 1804

Serving in England as a representative for the colony of Pennsylvania in 1770, Benjamin Franklin was introduced to a plant from the Orient called a Chinese garavance. Intrigued by the soft “cheese” that could be made from compressing the soaked garavance seeds, he sent samples on a ship heading back to the colonies. This was not the first instance of soybean [Glycine max (L.) Merr.] arriving on the shores of America; it had been quietly introduced into the colonies as early as 1765 when Samuel Bowen, a former sailor who had spent time in China, asked the Surveyor-General of Georgia to plant seeds of a “vetch” he had collected from his journeys (Hymowitz and Harlan, 1983; Hymowitz and Shurtleff, 2005). It is easy to take for granted today that the exotic fruits and vegetables, and even the grains and legumes, which we encounter every time we visit the local grocery store, have always been here. However, the vast majority of these plant products were foreign to the citizens of this young country only one or two centuries ago. As an illustration of this point, not one of the 20 crops considered most important as a food source originated in the USA (Raeburn, 1995). The crops we depend on today, such as maize (Zea mays), rice (Oryza sativa), cotton (Gossypium L.), and soybean, were introduced into this country, some repeatedly in different places and over different years, some glamorously as by Benjamin Franklin or Thomas Jefferson, and some so ubiquitously that they defined the entire
economy of regions of the country. In the South, fields of “King Cotton” stretched across the land, while north of this region lay the domain of another major crop, the “Corn Belt”.

Although considered one of the major food crops, soybean does not necessarily share the distinguished research history of other major crops such as maize and wheat (*Triticum aestivum*). It was not the subject of ground-breaking cytogenetics work, nor did it touch off the green revolution in agriculture in the 1960s. Soybean instead has the thankless task of providing the major component of “vegetable” oil, and is the staple of meatless alternatives in health food stores and elementary school cafeterias nationwide. In recent years, however, soybean has been launched into the spotlight as a health food that is a low-fat source of protein and can lower cholesterol. The U.S. Food and Drug Administration approved a recommendation in 1999 that Americans incorporate as much as 25 g of soy protein into their daily diets (Hasler, 2002). It is also the primary choice for a protein supplement in poultry and cattle feed. The USA is the world’s leading producer of soybean, producing upwards of 85 million metric tons in 2005 (American Soybean Association, 2005) and worth about $18 billion per year. Soybean is the subject of extensive genetics research as well. Thanks to the efforts of such scientists as Dr. Perry Cregan and Dr. Randy Shoemaker, the soybean genetic linkage map is saturated with DNA markers. Dr. Cregan is also currently incorporating a promising new type of marker that detects single nucleotide polymorphisms (SNPs) (Zhu et al., 2003). Soybean has over 340,000 expressed sequence tagged (EST) resources (http://www.plantgdb.org/), and a physical map of the soybean genome is on the verge of completion (Wu et al., 2004). In January 2006, the U.S. Department of Energy recognized the value of this crop, particularly for its potential as a biodiesel fuel, by announcing their decision to sequence its genome. Soybean cultivation as we
know it is thus the result of a rich history of plant breeding research, and all aspects of the crop are actively studied in numerous laboratories across the USA.

As far as evolutionary studies have discerned, cultivated soybean was domesticated from its wild relative *Glycine soja* in northeast China. Both species can easily hybridize and produce viable progeny, and cytogenetic studies also support the close relationship (Singh and Hymowitz, 1988). A careful historical analysis suggested the timeframe of domestication as during the Chou Dynasty, or approximately the 11th century B.C. (Hymowitz, 1970). From there, the soybean disseminated to central and southern China and Korea by the first century A.D., a route which corresponds to the expansion of the Chinese dynasties during this time (Hymowitz and Kaizuma, 1981). A study of chloroplast haplotypes of wild and cultivated soybean in China suggests that the soybean was independently domesticated multiple times, possibly from hybrid swarms of wild and cultivated forms (Xu et al., 2002). This repeated gene flow between wild and cultivated soybean plants underscores the diversity of soybean germplasm from Asia, and is a resource kept partially intact today through landraces grown in various regions. Although northeastern China is considered the center of origin for soybean, recent analyses of Asian soybean accessions using DNA markers distinguished Japanese accessions from Chinese and Korean germplasm, suggesting that this pool may also be a source of unique alleles (Abe et al., 2003; Ude et al., 2003). Various studies have noted the huge contribution of Chinese germplasm to the elite modern cultivars grown today in the USA (Li et al., 2001), and estimated that most of the diversity present in the Asian lines has not yet been utilized in U.S. cultivars (Abe et al., 2003). Through such efforts as a major soybean collection expedition to China, Japan, and Korea by the early 20th-century pioneers Howard Dorsett and Bill Morse (Hymowitz, 1984), the USDA now has a current holding of over 16,900 *G. max* and *G. soja* accessions or plant
introductions (PIs) (Singh and Hymowitz, 1999), with approximately 85% of the collection coming from the countries listed above, as well as from Russia (Vodkin, 1996). Countries such as Australia, Brazil, China, Germany, India, Indonesia, Japan, Russia, and South Korea also maintain significant soybean collections, resulting in more than 100,000 G. max accessions and 10,000 G. soja accessions (Vodkin, 1996). These collections are a literal goldmine, offering unexplored potential for improving such traits as yield, protein, and oil, or incorporating new alleles for disease and pest resistance.

The vast store of diversity represented by the plant introductions is critical for soybean. The majority of modern soybean cultivars grown in the USA derive their ancestry from fewer than 15 introduced lines, and cultivars released in the period from 1983 to 1988 shared 50% more genes in common with each other than did soybean cultivars released prior to 1954 (Gizlice et al., 1993). Collectively, the accessions from Asia or from other soybean germplasm pools are genetically diverse from U.S. elite lines (Li et al., 2001; Narvel et al., 2000). A source of diversity in the U.S. gene pool is additionally found in the distinct separation of northern from southern cultivars based on results from genetic marker studies (Kisha et al., 1998; Li et al., 2001; Sneller, 1994). These northern and southern U.S. cultivars represent unique gene pools, and evidence indicates that, due to the widespread use of relatively few ancestral lines such as ‘S-100’ and ‘CNS’, the southern gene pool is less diverse than the northern gene pool (Delannay et al., 1983; Gizlice et al., 1993). In a study of 122 northern and southern U.S. lines, the southern lines derived 73.2% of their parentage from the six most commonly used ancestral lines, while the northern lines derived only 59.7% of their parentage from their six most common ancestors (Sneller, 1994). Another finding estimated that 80% of the alleles contributed to 48 southern cultivars was derived from only seven plant introductions (PIs) (Delannay et al., 1983).
Based on these results, many researchers recommend crosses between northern and southern cultivars to incorporate new alleles into the separate breeding programs. Breeders have been reluctant to adopt this strategy, mainly because of the difficulty of working with lines that so vastly differ in time to maturity. More and more researchers have turned instead to “mining” the Asian germplasm of wild soybean accessions and PIs for new alleles.

Germplasm mining is not an altogether new enterprise. At least 23 nontimber crops in North America owe approximately 1% of their yield to genes derived from wild relatives (Raeburn, 1995). Increasing restrictions on pesticides place added pressure on breeding programs attempting to develop pest-resistant and competitive high-yielding lines. Novel alleles from germplasm can provide an environmentally safe and cost-effective alternative to chemical control; their use has proven most effective thus far as a source of resistance alleles to such diseases and pathogens as soybean rust (*Phakopsora pachyrhizi* Sydow), brown spot (*Septoria glycines* Hemmi.), powdery mildew (*Microsphaera diffusa* Cke. & Pk.), phytophthora root rot (*Phytophthora sojae* H.J. Kaufmann & J.W. Gerdemann), yellow mosaic virus, and soybean cyst nematode (*Heterodora glycines* Ichinohe) (Singh and Hymowitz, 1999). In addition, studies were initiated to evaluate hundreds of PIs for root-knot nematode (RKN; *Meloidogyne* spp.) resistance (Harris et al., 2003; Hussey et al., 1991; Luzzi et al., 1987) in order to first identify, and then transfer, key resistance alleles from the PI sources into agronomically acceptable breeding lines (Luzzi et al., 1996a; Luzzi et al., 1996b; Luzzi et al., 1997). The mining of alleles has now extended beyond disease resistance into the arena of tolerance to salt and herbicides, as well as increased seed yield (Concibido et al., 2003; Singh and Hymowitz, 1999). Researchers are also actively looking for genes that confer desirable fatty acid and amino acid profiles, among other traits (Vodkin, 1996). The discovery of novel alleles for high protein content has
been reported from wild soybean accessions, and a major quantitative trait locus (QTL) on linkage group I (LG-I) for high protein from *G. soja* has been confirmed in numerous studies (for a review see Chung et al., 2003). It appears that this high-protein allele is also present in ‘Danbaekkong’, a high-protein cultivar developed in South Korea and incorporated into the soybean breeding program at the University of Georgia (Harris, 2001).

Once the germplasm sources conditioning a desirable response in a trait have been identified, a lengthy process ensues in which the alleles are compared to those that have been previously identified to verify whether they are unique. If the alleles are unique, mapping studies are initiated to identify both the position of the allele in the species’ genome and the markers that are most closely associated with the allele. A close marker-allele linkage facilitates the introgression of the allele into breeding lines, which may ultimately be developed into cultivars. Before cultivars are deployed with this introgressed allele, however, it is critical to know its general effect on agronomic qualities of the crop, particularly on yield. The research presented in this dissertation demonstrates this entire process through the evaluation of novel alleles for two different traits, peanut RKN [*M. arenaria* (Neal) Chitwood or Ma] resistance and seed protein content, in soybean. The objectives of the dissertation were: 1) to determine the uniqueness and genomic positions of Ma resistance genes in three plant introductions from China, and 2) to evaluate the effect of a high-protein allele from a Korean cultivar on yield, protein and oil contents, amino acid composition, and other agronomic traits in the backgrounds of three elite soybean cultivars.

**References**


CHAPTER 2
REVIEW OF LITERATURE

Breeding for Resistance to Peanut Root-Knot Nematodes

Root-knot nematodes (RKN, *Meloidogyne* spp.) are a category of plant-parasitic nematodes especially adapted to the production environment in which crops such as soybean are grown. The parthenogenetic mode of reproduction, the large number of eggs produced, the ability to overwinter on weeds during fallow periods, and the facilitation of disease complexes all compound the damage to a soybean crop. Damage from all categories of plant-parasitic nematodes reduced the annual yield of soybean by an estimated 11 to 15 million kg in the USA from the years 1996 to 1998 (Wrather et al., 2001a), and caused growers in the southern region an economic loss of roughly $79.5 million over a period from 1994 to 1996 (Pratt and Wrather, 1998). The four most common RKN species, *Meloidogyne incognita* (Kofoid and White) Chitwood (Mi, southern RKN), *M. arenaria* (Neal) Chitwood (Ma, peanut RKN), *M. javanica* (Treub) Chitwood (Mj, Javanese RKN), and *M. hapla* Chitwood (Mh, northern RKN), are the cause of 95% of incidences of root-knot nematode parasitism on crops worldwide (Hussey and Janssen, 2002). Of the four species, only Ma and Mi are of most concern to soybean growers in the southeastern USA (Boerma and Hussey, 1992; Rodríguez-Kábana et al., 1991). More effort, however, has been focused on obtaining high levels of resistance to Mi. As a possible consequence of this focus, the predominant culture of Mi-resistant cultivars may have increased damage to crops from Ma (Fassuliotis, 1987; Garcia and Rich, 1985; Ibrahim and Lewis, 1986; Ibrahim and Lewis, 1993). Of the two races of Ma that can infect soybean, race 1 is a persistent problem because its major host, peanut (*Arachis hypogaea* L.), is grown on substantial
hectareage in the southeastern USA and may be rotated with soybean (Rodríguez-Kábana et al., 1991). Both races, however, are increasing in population density in this region, and race 2 appears to be more pathogenic on soybean than race 1 (Pedrosa et al., 1994). Race 2 is also more widely distributed, making up 84% of worldwide Ma populations (Riggs, 1991). Yet as a result of environmental and financial concerns about the use of nematicides, there are limited means of chemical control for these nematodes. Resistant cultivars and crop rotations are often the only tools available to soybean producers to reduce nematode damage (Boerma and Hussey, 1992). Highly resistant cultivars are particularly attractive since these cultivars can significantly reduce nematode levels in the soil, creating a better soil environment for growing less resistant but profitable crops in following years. Since the level of Ma resistance in currently grown soybean cultivars is not adequate to prevent yield loss, previously described sources of Ma resistance require better characterization and new sources of Ma resistance need to be evaluated and their resistance genes mapped.

**Plant-Parasitic Nematodes**

Although the word “nematode” is not widespread in the public vernacular, these unsegmented roundworms are ubiquitous. A common anecdote is that if you removed all the soil and water from the earth and looked at the result from space, the vast number of nematodes that live in the earth and water would still maintain the outline of the planet. The phylum Nematoda is considered the most successful of metazoans, representing 80 to 90% of all animals (Bird and Koltaï, 2000). The nematode *Caenorhabditis elegans* is a model organism that has tremendously benefited both medical research and in understanding processes such as gene silencing and aging (Davis, 2004). However, nematodes are most notorious for the diseases they cause in humans and animals worldwide, such as elephantiasis, river-blindness, and heartworm (Bird and
Kaloshian, 2003). The plant-parasitic nematodes are an equally devastating subset of nematodes, globally causing billions of dollars in crop yield loss, or an equivalent of 12.3% yield loss in the major crops that are grown worldwide (Sasser and Freckman, 1987).

Plant-parasitic nematodes are capable of infecting most of the 250,000 described flowering plants, and can occupy all organs of vascular plants, from root tips to shoot meristems (Bird and Koltai, 2000). These roundworms are generally 1 mm long, and are thus barely visible to the naked eye. As obligate parasites, they depend on the cytoplasm of plant cells for survival (Williamson and Hussey, 1996). Crop plants grown in tropical or temperate regions are especially susceptible, as these are areas where most plant-parasitic nematodes flourish.

Nematodes are easily distributed by agriculture through transplants, cuttings or tubers, and the movement of soil between fields on machinery (Trudgill and Blok, 2001). This facilitates the introduction of nematodes to potential new hosts, plants that may not have had previous exposure to nematodes and will thus have no developed resistance (Blok et al., 1997). Nematodes will invade these plants, and establish a host-pathogen interaction that is more sophisticated than that of most plant-parasitic viruses, bacteria, and fungi (Bird and Koltai, 2000). A critical feature that enables this parasitism is a hollow stylet that allows the nematode to penetrate, feed upon, and inject secretions into plant tissue (Trudgill, 1991). This is also a key characteristic in the taxa of the division Secernentea to which many plant-parasitic nematodes, including the root-knot and cyst nematodes, belong. Within this division, the Aphelenchida taxon is distinguished by slender stylets with few variations, while the Tylenchida taxon is characterized by a diversity of sizes and shapes of the stylet (Clark, 1994). Another feature is the modification of the esophageal glands to produce secretions directing specific aspects of the parasitism, a feature discussed in more detail later.
The plant-parasitic nematodes are also categorized by the movement of the nematode, whether inside or outside of the root. The ectoparasites do not physically enter the root, instead feeding on cells close to the exterior or, more rarely, having a long stylet that allows feeding in deeper tissues. The ectoparasites may be migratory or sedentary, terms referring to the activity of the nematode in moving about the root to establish new feeding sites. In contrast, the endoparasites enter the root, traveling through the symplast or apoplast of the plant tissue. While the migratory endoparasites feed on plant cells as they move through the tissue, the sedentary endoparasites travel to a specific region of the plant tissue, such as the vascular cylinder, and direct the formation of a complex multicellular feeding site. The sedentary endoparasites require specialized feeding sites to develop into mature adults, which most importantly include nonmobile females that produce vast quantities of eggs (Trudgill, 1991). These migratory and sedentary endoparasites are the most damaging of the plant-parasitic nematodes (Bird and Koltai, 2000), and of these the sedentary endoparasites of the Heteroderidae family, the cyst and RKNs, are particularly devastating to agriculture (Williamson and Hussey, 1996).

**Cyst Nematodes**

Cyst nematodes have narrow host ranges compared to the RKNs. This group of nematodes will invade the vascular cylinder, organizing the incorporation of up to 200 surrounding cells into a feeding site called a syncytium. Only the female cyst nematode can establish this feeding site, and the subsequent swelling of her body as she feeds and molts will eventually rupture the root. Her body ultimately hardens into a protective cyst around the eggs that occupy the interior cavity. The soybean cyst nematode (SCN, *Heterodora glycines*) is currently the most damaging pathogen to U.S. soybean production, resulting in an annual loss of 7.59 million metric tons in yield (Wrather et al., 2001b). Although SCN is the most important
pest on soybean, RKNs cause more damage to the global food supply than any other category of plant-parasitic nematodes (Bernard, 1989). The wide host range of RKNs means that any crop grown in rotation with soybean is probably susceptible as well. Most of the major crops grown in the southeastern USA are indeed susceptible to RKNs (Luzzi, 1986). As discussed in more detail below, this limits growers’ alternatives in managing RKN infestations in the field.

**Root-Knot Nematodes**

The RKNs belong to the genus *Meloidogyne*, which contains over 54 described species (Trudgill and Blok, 2001). *Meloidogyne* spp. can parasitize many types of plants, from food crops to ornamental species, that grow in both tropical and temperate climates (Eisenback and Triantophyllou, 1991). All RKNs cause the formation of feeding sites from giant polyploid cells that have undergone multiple cycles of DNA replication without cytokinesis. The formation of giant cells in the vascular tissue is highly beneficial to the nematode, as they provide access to the nutrients and water from the phloem and xylem tissue (Eisenback and Triantophyllou, 1991). The complex interaction of nematodes with their hosts lasts more than a month, and involves a sophisticated interplay of signals and responses from both the nematode and the host (Williamson and Hussey, 1996). The subventral and dorsal gland cells in the esophagus of both SCNs and RKNs produce secretions that are involved in the invasion process and the development and maintenance of the feeding site (Hussey et al., 2002). Some of the secretions are cell-wall degrading enzymes (CWDE) such as cellulase, pectate lyase, and polygalacturonase, which can facilitate RKN movement through the root by softening the cell walls (Bird and Kaloshian, 2003). Other secretions result in the formation of a syncytium of four to five multinucleate giant cells in the vascular cylinder, among which the female alternates her feeding. These giant cells have the characteristics of reduced plasmodesmatal connections and
modifications to the cell wall that increase the surface area of the plasma membrane (Bird and Koltai, 2000).

The plant responds to this invasion by producing damaging reactive oxygen species and signaling molecules such as lipid fragments, ethylene, and salicylic acid (Blok et al., 1997). Plant genes such as peroxidases and chitinases are up-regulated, the cells around the invasion site show an altered balance of auxin and cytokinin (Bird, 2004), and callose is deposited as a wound response in the regions where the nematode has inserted its stylet (Williamson and Hussey, 1996). In a compatible reaction, a scenario in which the susceptible plant allows syncytium formation, the RKN female undergoes progressive molts accompanied by body swelling, a series of events similar to that in female cyst nematodes. The cortical and pericycle cells around the feeding site expand and divide rapidly to form a knot in the root called a gall (Bird and Kaloshian, 2003). Galls typically form only 1 or 2 d after nematode penetration, and the gall size normally relates to the number of nematodes in that particular root area (Eisenback and Triantophyllou, 1991). Most of the economically important Meloidogyne spp. are mitotically parthenogenetic: Though RKN males may form under specific stress conditions, they may or may not mate with females. The RKN female does not harbor eggs within the body, instead depositing them in an egg sac directly outside the body. Though some Meloidogyne spp. have limited host ranges similar to that of the cyst nematodes, the economically important species Ma, Mi, and Mj are able to invade over 2000 plant species, with species representing almost every plant family (Roberts, 1995; Trudgill and Blok, 2001).

Many studies contend that the mitotic parthenogenetic or apomictic reproduction of the RKN species is a strategy that prevents coevolution with a specific host. Apomixis is a method that maintains favorable gene combinations, and allows RKNs that are dispersed to new
environments to reproduce rapidly and without a partner (Blok et al., 1997). Other characteristics of RKNs that contribute to their wide host range include the ability to hatch without a specific signal from a nearby plant, and polyploidy, which allows rapid accumulation of gene products (Trudgill and Blok, 2001).

The wide host range and the fecundity of RKNs present difficulties in managing nematode infestations (Roberts, 1995). The wide dispersal of Mi, in particular, has led to its distinction as possibly “the single most damaging crop pathogen in the world” (Trudgill and Blok, 2001). It is also the nematode species most frequently found in soybean fields in Georgia (Herman et al., 1991). Ma is found within the same latitudinal distribution as Mi, but it is not as common as either Mi or Mj. However, its biochemistry and cytological profile are more variable and atypical than either Mi or Mj (Eisenback and Triantophyllou, 1991). Further study of this nematode’s interaction with its plant hosts may yield different insights than what is already known about either of the more common RKN species.

Methods of Nematode Control

Three common methods to manage nematode populations are chemical nematicides, crop rotations of hosts with nonhost plants, and resistant cultivars. Though chemical nematicides are an effective means of control, they are expensive to apply and can cause damage to non-target organisms (Trudgill, 1991). Their expense cannot be justified on low-value crops such as soybean, and many of the most effective chemicals have already been banned from the market. Crop rotations are frequently used to improve soil quality, such as a rotation of a grass crop with a nitrogen-enriching legume, or to control diseases by the rotation of resistant cultivars with susceptible cultivars or host crops with nonhosts. Crop rotations have been used successfully in nematode control: A rotation of susceptible cotton cultivars with resistant soybean and corn
stabilized cotton yield in fields infested with the reniform nematode (*Rotylenchulus reniformis*) (Davis et al., 2003). However, the wide host range of RKNs means that in many cases nonhosts may be unavailable or not economically acceptable (Herman et al., 1991; Luzzi, 1986). Even if a rotation scheme is possible, it may only control one species of nematode; thus the presence of multiple species may still damage the crop or result in higher nematode population densities (Weaver et al., 1988). The use of resistant cultivars with specific genes that reduce the damage or reproduction of invading nematodes may be the most economical solution to nematode control (Niblack et al., 1986). Resistant cultivars require little specialized equipment or technology, can shorten the length of each crop rotation, and do not pollute the environment (Trudgill, 1991). SCN-resistant near-isogenic lines (NILs) out-yielded their susceptible counterparts in SCN-infested fields (Brucker et al., 2005). Cultivar resistance can also remain effective for several decades since RKNs have limited gene flow due to their parthenogenetic mode of reproduction and their limited mobility in the soil (Blok et al., 1997). As an indication of the predominance of each of these methods of nematode control, a study in Florida found that soybean growers were using Mi-resistant cultivars in 70% of their fields, crop rotations in 35% of the fields, and nematicides in 9% of their fields (Garcia and Rich, 1985).

Certain caveats exist for the use of nematode-resistant cultivars. Resistance has not yet been identified in certain plant species, or is available for only certain species of plant-parasitic nematodes. For example, there are few genes conferring resistance to the RKNs within the cucurbits such as squash and pumpkin (*Cucurbita* spp.) (Fassuliotis, 1987). Furthermore, some resistance factors are only effective against certain races of the same nematode species (Williamson and Hussey, 1996), leaving crop plants completely susceptible to other races. In one study, expression of a resistance gene to SCN in the absence of the pathogen has been
associated with significantly reduced soybean yields (Kopisch-Obuch et al., 2005). However, these deficiencies have become easier to ameliorate because of advances in the introgression of novel genes or improved access to germplasm sources with new resistance genes. Cultivar resistance has mainly been identified for those nematode species that establish more sophisticated relationships with their hosts, which usually involves the development of a feeding site. This includes the sedentary endoparasites, some of the migratory endoparasites, and even a few of the ectoparasites like *Xiphinema* spp. that can feed deep within root tips (Blok et al., 1997; Roberts, 1992). Most of the nematode resistance characterized in crops is specific against the RKNs and cyst nematodes.

**Terminology to Describe the Plant-Nematode Interaction**

Nematode resistance is typically conditioned by major genes, meaning that the resistance acts qualitatively and is likely conferred by a few dominant, large-effect genes. However, SCN resistance genes in soybean are recessive (Roberts, 1992). Additionally, resistance to the potato cyst nematode *Globodera pallida* is quantitative, or conditioned by several small-effect genes. Nematode resistance is typically manifested as decreased or delayed reproduction relative to that on the susceptible plant (Niblack et al., 1986; Roberts, 1992). Over time, this may decrease nematode populations in the soil (Herman et al., 1991). For soybean, resistance to RKNs is seen as a reduction in the number of galls and eggs produced on the root system. This definition of resistance is distinctly different from the effect of tolerance, which is the ability of the plant to withstand the nematode feeding without a corresponding decrease in yield. A plant breeder ideally would like to develop cultivars that are both tolerant and resistant, since this would both decrease nematode reproduction (the resistance effect) and maintain the yield at an acceptable level (the tolerance effect).
Host and Nonhost Resistance

Resistance can be further described as host or nonhost types of resistance. Nonhost resistance is defined as a reduction in nematode entry into the root (Trudgill and Blok, 2001). A nonhost is thus a plant that lacks some component(s) the nematode requires to invade the plant and establish a compatible interaction. Host resistance, in contrast, is seen as a decrease in nematode reproduction that is conditioned by one or two major-effect genes. Some 40 major effect resistance genes have been cloned in various plants in the past few years, and many details are known about their activity and arrangement in the genome (Martin et al., 2003). Most characterized resistance (R) genes have distinct motifs that readily identify them as resistance genes, such as a leucine-rich repeat (LRR) region, a nucleotide-binding site (NBS) motif, or a coiled-coil (CC) motif (Martin et al., 2003). Many resistance genes seem to function through a gene-for-gene interaction with the pathogen. A resistant plant, therefore, is able to recognize some gene product that the pathogen secretes, and activates a specific defense response in return. Many of the events of the defense response are similar to those activated in the susceptible response mentioned above, such as the production of reactive oxygen species and signals such as lipid fragments, ethylene, and salicylic acid. However, the timing and extent of this signal production is more rapid and extensive in resistant plants. Transcription factors are activated, specific defense genes are expressed, and the hypersensitive response (HR) results (Blok et al., 1997). The HR is a form of programmed cell death that is actively triggered by a plant that is resistant to an invading pathogen. The secreted pathogen product that the plant recognizes is called an avirulence (avr) protein and the isolate or biotype of the pathogen that secreted the avr protein is also called avirulent. Although no avr genes have yet been isolated from nematodes,
certain nematode isolates or strains are described as virulent or avirulent to different nematode resistance genes.

In many crops such as soybean, abundant numbers of QTLs conditioning resistance have been identified, most notably for this study from three species of root-knot nematodes (Tamulonis et al., 1997a; Tamulonis et al., 1997b; Tamulonis et al., 1997c). However, the putative resistance gene(s) accounting for the QTL have not been isolated, although the cloning of both minor and major QTLs is becoming more common (Paran and Zamir, 2003).

**Pathogen Terminology**

Populations of nematodes may vary in host range and in pathogenicity, and these variations are categorized as host races (Trudgill and Blok, 2001). Host races have been defined for Ma and Mi, but are not as clearly distinguished for Mj and Mh (Hussey and Janssen, 2002). The definition of race differs between types of nematodes; while SCN races are defined by different cultivars and lines of the same species (i.e., soybean), RKN races are defined by different host species (Riggs, 1991). For example, two races of Ma can attack soybean, but only race 1 can parasitize both soybean and peanut (Rodriquez-Kábana et al., 1991). Other than this differential host reaction, this is the only known difference between the two races. Therefore, both races differ in pathogenicity as well as host range. Pathogenicity is the general ability to cause disease or damage. The term virulence, in contrast, reflects the capacity of the pathogen to overcome or suppress specific R genes. Virulence was mentioned above in reference to avirulent isolates and avirulent proteins. Biotypes or pathotypes are often used in lieu of races to categorize nematode populations as virulent or avirulent (Roberts, 1995). Virulent biotypes may be selected *in vitro* from formerly avirulent biotypes or may naturally occur in different populations (Roberts, 1995). To date, no avirulence proteins have been isolated from nematodes.
Avirulence products isolated from other plant pathogens are diverse molecules, ranging from coat proteins in viruses to a metalloprotease in the rice blast fungus *Magnaporthe grisea* (Martin et al., 2003).

**Increasing Threat of Ma to Soybean Production**

The narrow genetic base of elite soybean cultivars precludes significant variation for many traits, including RKN resistance. One statistic estimates that almost 80% of the diversity in elite U.S. soybean cultivars was contributed by fewer than 10 plant introductions (PIs) (Delannay et al., 1983). Yet multiple resistance genes are needed in cultivars to provide more durable resistance, particularly since soybean fields in the southeastern USA may be infested with different species of plant-parasitic nematodes (Niblack et al., 1986). *Meloidogyne arenaria* (Ma), or the peanut root-knot nematode, is steadily increasing in population density in the southeastern USA, consists of many isolates of two races, and can cause yield reductions on susceptible cultivars (Kinloch et al., 1987; Rodríguez-Kábana and Williams, 1981). Race 2 of Ma does not parasitize peanut, but it appears to cause greater yield reduction on soybean than race 1 (Pedrosa et al., 1994). The yield reduction was most likely a result of the greater ability of race 2 to form galls and egg masses on soybean roots than that of race 1 (Ibrahim and Lewis, 1993). This effect, however is more likely to depend on the particular isolate used rather than the particular race, as another study reported that the level of galling between Ma-resistant and Ma-susceptible cultivars did not appear to greatly differ when either race was studied (Koenning and Barker, 1992). Studies of different race 2 isolates have found significant differences in aggressiveness among them and in comparison to race 1 isolates (Carpenter and Lewis, 1991b; Hiatt III et al., 1988; Ibrahim and Lewis, 1993).
In the context of this study, Ma poses an increasing threat to soybean production for the following reasons: i) much attention has been focused on obtaining cultivars resistant to Mi, at the expense of increasing Ma populations in many states in the southeastern USA (Fassuliotis, 1987; Fortnum et al., 1984; Garcia and Rich, 1985; Ibrahim and Lewis, 1993; Rodríguez-Kábana and Williams, 1981; Schmitt and Barker, 1988), ii) peanut is a major host of race 1 of Ma, and soybean is a rotation crop with peanut (Rodríguez-Kábana et al., 1991), iii) Ma can reproduce better at lower temperatures than Mi, presenting a risk as the practice of planting earlier maturing soybean varieties, which have not traditionally been developed with RKN resistance, in the southeastern USA becomes increasingly popular (Ibrahim and Lewis, 1986; Kirkpatrick and May, 1989), iv) Ma tends to be more aggressive on soybean than Mi, posing greater damage to soybean as its populations increase (Starr, 1989), v) many nematicides are not as effective against Ma as against Mi (Barker et al., 1981; Ibrahim and Lewis, 1993), and vi) even if Ma populations remain low, crops growing in fields that are infected with more than one species of nematode suffered growth inhibition, even though the cultivar was tolerant to one of the nematode species, by the synergistic effect of both nematode species (Ibrahim and Lewis, 1986). Furthermore, Ma populations are expected to continue increasing in light of continued cultivation of Mi-resistant cultivars (Rodríguez-Kábana and Thurlow, 1980).

**Previous Attempts to Characterize and Map Ma Resistance Genes in Soybean**

Most soybean cultivars exposed to Ma do not possess adequate resistance to avoid yield loss (Kinloch et al., 1990). A comprehensive study of 2370 PIs identified PI 200538 and PI 230977 as highly resistant to both Ma (races 1 and 2) and Mj (Luzzi et al., 1987). The level of Ma resistance found in both PI sources was higher than that in the resistant check cultivar Jackson. While neither PI differed in gall index and Ma reproduction in the greenhouse, PI
200538 was less damaged by Ma in the field than was PI 230977 (Pedrosa et al., 1994). Quantitative trait locus (QTL) mapping studies and introgression of Ma resistance into breeding lines have since proceeded solely with PI 200538. PI 230977, in contrast, has been the focus of studies on resistance to Mj. Another more recent study established PI 594403, PI 594427C, and PI 594651L2 as promising sources of resistance to Ma (Harris et al., 2003). Although these new PIs had more galls than PI 200538, both PI 594403 and PI 594427C had fewer or equal numbers of eggs, respectively. While the genes conditioning the nematode resistance in PI 200538 and PI 230977 have been transferred into the respective breeding lines G93-9106 and G93-9223 (Luzzi et al., 1996; Luzzi et al., 1997), the resistance in PI 594403, PI 594427C, and PI 594651L2 has not yet been fully exploited.

Several crosses were evaluated to determine if the resistance genes in PI 200538, PI 230977, and Jackson were unique. The wide segregation of resistance in both the F2 individuals and the F3 families from each cross suggested that Jackson and the two PIs did not share all the same resistance genes (Luzzi et al., 1995a). Mapping studies further evaluated F2 populations of PI 200538 x CNS (a cultivar that is Ma-susceptible) for Ma resistance and PI 230977 x CNS for Mj resistance. These studies established that both PIs had a major resistance QTL on LG-F, but had minor QTLs on different linkage groups (Tamulonis et al., 1997b; Tamulonis et al., 1997c). While the large-effect resistance QTL may confer more broad resistance, as to both Ma and Mj, minor QTLs are typically thought to be modifiers that may condition greater specificity to a particular species (Lee et al., 1999). The major QTL on LG-F and the minor QTL on LG-E in PI 200538 accounted for 51% of the variation in gall number caused by Ma in the F2:3 population, while the major QTL on LG-F and the minor QTL on LG-D1 in PI 230977 explained 54% of the variation in gall number caused by Mj.
The Resistance Gene Cluster on LG-F

The area of LG-F to which the major QTL in PI 200538 and PI 230977 mapped is, along with LG-J, a region that contains clusters of resistance genes (Graham et al., 2000; Jeong et al., 2001; Kanazin et al., 1996; Yu et al., 1996). A study of soybean unigene sets, which are developed from clustered ESTs, found as many as 326 unigenes that could be tentatively classified as resistance gene candidates based on sequence similarity to known resistance genes (Tian et al., 2004), revealing the extent of resistance genes that are possibly present in these and other clusters of resistance genes in soybean. Clusters of resistance genes have also been reported in rice (*Oryza sativa* L., Ramalingam et al., 2003), maize (*Zea mays*, Quint et al., 2003), potato (Trognitz et al., 2002), lettuce (Meyers et al., 1998), tomato (Dickinson et al., 1993), and *Medicago truncatula* (Tian et al., 2004), all of which, like soybean, are known or suspected to have been polyploid at some point in their evolutionary history. Resistance genes are thought to have arisen and evolved through duplication of chromosome segments or of a single locus, and the resulting R genes may then be altered or re-distributed by recombination, homogenization, mutation, selection, or additional duplication events (Leister, 2004; Richter and Ronald, 2000).

The resistance gene cluster on LG-F in soybean is located in the interval between the RFLP markers B212 and K644, which is upstream of R045, and is known to contain resistance genes to a wide range of pathogens, from the RKN spp. Ma and Mj (Tamulonis et al., 1997b; 1997c) to viruses such as soybean mosaic virus (Yu et al., 1994), peanut mottle virus (Saghai Maroof, unpublished), and tobacco ringspot virus (Fasoula et al., 2003), the oomycete fungus *Phytophthora sojae* (Diers et al., 1992), the bacteria *Pseudomonas syringae* pv. *glycinea* (Ashfield et al., 1998), and the insect *Helicoverpa zea* Boddie (Rector et al., 1999). This region of the genome is not only rich in resistance genes, but also in the number of alleles conferring different
specificities at each R gene locus (Ashfield et al., 1998). Similarities between resistance gene-like sequences from this area of LG-F and sequences from common bean suggest that this cluster of R genes may have been present in the ancestor of both common bean and soybean (Ashfield et al., 2003). Sequence analysis of resistance-like gene clones isolated from this region of LG-F revealed two major classes of clones out of a total of 11 classes isolated from the soybean genome. While class b clones are more similar to members of the TIR-NBS-LRR subclass such as the N and L genes from tobacco (Nicotiana tabacum) and flax (Linum usitatissimum), respectively, and also include the resistance gene candidate L20a that is used extensively in this study (Hayes et al., 2000), the class j clones are most similar to CC-NBS-LRR genes such as Rps2 in Arabidopsis thaliana (Yu et al., 1996). The two classes of clones were separated only by approximately 0.30 cM, which emphasizes both the diversity and tight linkage of resistance genes in this region. However, the physical distance in this region has been roughly estimated between 225 to 450 kilobases (kb) per cM (Hayes et al., 2004), which means that the two types of clones are still separated by thousands of base pairs. The difficulties of mapping in a region that contains both tandem duplications and clusters of structurally similar genes as on LG-F was underscored in a similar study on LG-A2 of soybean (Lewers et al., 2002), and may explain why map-based cloning of R genes in soybeans has not been previously reported (Ashfield et al., 2003). Since the QTLs in PI 200538 and PI 230977 do map to a known cluster of R genes, it is not illogical to suspect that two tightly linked QTLs may separately condition resistance to Ma and Mj, or, conversely, to suppose a single QTL may confer the same response.

RFLP mapping in PI 200538 x CNS and PI 230977 x CNS populations in previous studies effectively identified significant QTLs for Ma and Mj resistance, particularly the major QTL on LG-F (Tamulonis et al., 1997b; 1997c). However, both mapping studies for Ma and Mj
resistance in PI 200538 and PI 230977 were completed prior to 1997, and several new marker
types have since become available to fine map the region around the identified QTLs. Simple
sequence repeat (SSR) markers target specific classes of repeats in the genome, and are
particularly useful in detecting polymorphisms in soybean (Akkaya et al., 1992). SSRs not only
detect more polymorphisms than RFLPs, AFLPs, or RAPDs (Powell et al., 1996; Smith et al.,
1997), but are codominant and are, with some exceptions, evenly spread across the genetic map
(Cregan et al., 1999). There are now over 1000 SSRs available for genetic mapping in soybean,
and an updated map was recently released that combines genetic maps from five different
soybean genotypes into a consensus map of marker orders and genetic distances (Song et al.,
2004). In addition, sequences have recently become available in the form of expressed sequence
tags (ESTs) and bacterial artificial chromosomes (BACs), providing a template to design primers
for mapping and other purposes detailed below.

Although SSRs have been mapped to either side of the two RFLPs B212 and K644 on
LG-F, there are few SSRs between them. A previous study noted that regions of the soybean
genome containing resistance gene clusters, particularly on LGs F and J, tended to have few SSR
markers in those intervals (Demirbas et al., 2001). This may be explained by limited repetitive
sequences, which SSRs often target, within such a gene-rich interval. However, several studies
have established new sources of molecular markers in this important region on LG-F. For
example, studies with the soybean mosaic virus in soybean led to the identification of a SCAR
(sequence characterized amplified region) marker that is linked to the gene on LG-F conditioning
resistance to this virus (Zheng et al., 2003). This SCAR marker, known as OPN11, was mapped
in a recombinant inbred line (RIL) population to within 0.7 cM of B212, and is extensively used
in the research reported in this study. Another source of markers is resistance gene analog
(RGA) clones that have been isolated from cDNA and genomic libraries of soybean (Jeong et al., 2001; Peñuela et al., 2002). Resistance gene analogs, also described as resistance gene candidates (RGCs) or resistance-like genes (RLGs), are tagged by sequence similarity to conserved regions of previously identified resistance genes, and have been isolated from soybean (Graham et al., 2000; Hayes et al., 2000; Yu et al., 1996), as well as numerous other plant species such as Arabidopsis (Michelmore, 2003), rice (Ramalingam et al., 2003), potato (Trognitz et al., 2002), and maize (Quint et al., 2003). Sequences from RGA clones on LG-F are deposited in GenBank (Hayes and Saghai Maroof, 2000; Hayes et al., 2004; Jeong et al., 2001; Peñuela et al., 2002), and may be used to isolate the resistance gene itself (Ashfield et al., 2004), or simply to develop PCR-based primers to further resolve the position of the QTL on LG-F. A fine-scale genetic map of the region around the B212 marker is available that lists the current SCAR and RGA-based markers that have been mapped to this area (Gore et al., 2002). Other legumes that have sequence resources, such as Lotus japonica and Medicago truncatula, may be used to find orthologous sequences in this area of LG-F. Interestingly, the former species has been appraised as an excellent model to study legume-nematode interactions (Lohar and Bird, 2003), and thus may prove to be a good resource for more than just sequence information. Many single nucleotide polymorphisms are currently being developed for soybean (Zhu et al., 2003). RGAs have been reported as holding great promise for the development of SNPs (Quint et al., 2002).

Identifying a tightly linked marker to these QTLs will result in more efficient genotyping in marker-assisted selection (MAS) and facilitate the introgression of all resistance alleles into a single breeding line. If a single QTL is found to condition Ma resistance, it may be the site of a rare allele that underlies all Ma resistance currently found in elite cultivars such as ‘Haskell’ or
‘Jackson’. The recurrent introduction of the same alleles from different germplasm sources is a
commonly reported phenomenon in the literature (Michelmore, 2003; Sebolt et al., 2000). In the
case of Mi, a simple sequence repeat marker (SSR) tagged a rare resistance allele on LG-O that
was found in almost all previously identified Mi-resistant cultivars (Ha et al., 2004).

Screening for Ma Resistance in the Greenhouse

Early studies that screened plants for nematode resistance used naturally infested fields or
inoculated plants in field microplots (Birchfield and Harville, 1984; Carpenter and Lewis, 1991a;
Herman et al., 1990; Niblack et al., 1986; Weaver et al., 1988). However, greenhouse screening
for nematode resistance gives results that are highly associated with those of field and microplot
screening, but offers the advantages of screening on a year-round basis and uniformity of
inoculation and growing conditions (Hussey and Boerma, 1981; Niblack et al., 1986). It allows
for plant screening with only one nematode species at a time, while a naturally infested field may
contain several different species (Niblack et al., 1986).

However, there are still issues to consider when designing root-knot nematode screening
experiments in the greenhouse. The experiment can be influenced by the particular environment
in which the soybeans are grown, the genetic composition of the lines being tested, the particular
race or isolate being used, and the guidelines used to distinguish resistant from susceptible plants
(Fassuliotis, 1987). The timing of the nematode screening will vary depending on the season or
the particular plant genotypes used (Saichuk et al., 1976). Screenings taking place in cooler
months will take longer than those in warmer months because the lower temperature slows down
the rate of nematode development and reproduction. Progeny testing from resistant x resistant
crosses requires higher levels of inoculum than resistant x susceptible crosses to fully
differentiate moderately from highly resistant lines (Hussey and Janssen, 2002; Luzzi et al.,
In contrast, progeny tests of resistant x susceptible crosses are normally inoculated with fewer eggs. The inoculum concentration, however, will vary greatly depending on the aggressiveness of the particular Ma race or isolate (Carpenter and Lewis, 1991a). If too many eggs are used, it becomes difficult to distinguish susceptible from resistant plants (Luzzi, 1986). Though galling indices are commonly used to score for resistance (Saichuk et al., 1976), gall formation and nematode reproduction can be controlled by different genes (Luzzi et al., 1987), and are often mutually exclusive of each other (Hussey and Boerma, 1981). In some cases, therefore, it is necessary to screen both. This is particularly true for those plant species where the extent of root galling and the density of egg mass production are unrelated (Hussey and Janssen, 2002). In soybean, however, root gall index and egg density were found to have a fairly good correlation of 0.6 to 0.7 (Luzzi, 1986), and most studies report the extent of galling as evidence of resistance or susceptibility. Several studies report a lack of association between the gall number of single F2 plants and their derived F3 families (Luzzi et al., 1995a; 1995b), and suggest characterizing resistance based on replicated families rather than individual F2 plants (Luzzi et al., 1994). In accordance, our experiments will confirm the level of Ma resistance expressed by individual F2 plants with replicated F2:3 families and F2:4 families.

Objectives

The literature review outlined several pertinent questions related to Ma resistance in soybean that remain to be addressed. From a broad view, there are at least three major research goals that can be formed. The first is to evaluate the recently identified PI 594403, PI 594427C, and PI 594651L2 as sources of novel Ma resistance genes that differ from those previously mapped in PI 200538. Resistant x resistant crosses of each of the new PIs to PI 200538 were therefore initiated and evaluated to determine whether the progeny significantly segregated in
their extent of galling in two different greenhouse experiments. If the information from the first objective indicates that some or all of the PIs seem to contain novel resistance genes, the second objective would be to map the novel resistance genes in Ma resistant x Ma susceptible crosses. A third and final objective would be to determine whether the resistance alleles on LG-F mapped in PI 200538 and PI 230977 populations, as well as in any of PI 594403, PI 594427C, and PI 594651L2, are present in modern Ma-resistant cultivars. The resolution of this question depends on obtaining a tight marker-QTL linkage through saturating this region of LG-F with additional markers mapped to this region.

References


CHAPTER 3

ADDITIONAL SOYBEAN RESISTANCE GENES TO PEANUT ROOT-KNOT NEMATODES*

*Yates, J.L., R.S. Hussey, S.L. Finnerty, and H.R. Boerma. To be submitted to Crop Sci..
Abstract

Peanut root-knot nematodes [Meloidogyne arenaria (Neal) Chitwood or Ma] are an increasingly common pest in the southern USA in which crops such as peanut (Arachis hypogaea L.) and soybean [Glycine max (L.) Merr.] are grown. Five soybean plant introductions (PIs) have been identified as highly Ma resistant. While the resistance alleles in PI 200538 have been characterized, the Ma resistance allele(s) in PI 230977, PI 594403, PI 594427C, and PI 594651L2 have not yet been identified. In order to determine whether these PIs contain different resistance alleles from those in PI 200538, populations developed from the PIs crossed to PI 200538 were evaluated for Ma response in the greenhouse. Both highly galled and highly resistant F2 plants were observed in each of the four crosses, and these plants were transplanted to obtain F2-derived progeny for two confirmation experiments. In the crosses PI 200538 x PI 594403, PI 200538 x PI 594427C, and PI 200538 x PI 594651L2, the highly galled F2 plants had progeny that sustained significantly greater galling (P<0.001) in both confirmation experiments than the progeny derived from highly resistant F2 plants. While the F2-derived progeny from PI 200538 x PI 594403 had more variability than was observed in the other crosses, the means of the highly galled and highly resistant lines were significantly different. Only the F2-derived progeny of PI 200538 x PI 230977 did not perform similarly to the original F2 plants (P>0.4 in both confirmation studies). Single lines within all of the populations were significantly different from each other or from the parents. Based on either the means of the gall class differences or the single-line differences, there is evidence that PI 230977, PI 594403, PI 594427C, and PI 594651L2 contain unique resistance genes that, when combined with PI 200538-derived Ma resistance, could improve the level of Ma resistance in southern soybean cultivars.
Introduction

The RKNs belong to the genus *Meloidogyne*, which includes over 54 described species (Trudgill and Blok, 2001). The four most common RKN species, *Meloidogyne incognita* (Kofoid and White) Chitwood (Mi, southern RKN), *M. arenaria* (Neal) Chitwood (Ma, peanut RKN), *M. javanica* (Treub) Chitwood (Mj, Javanese RKN), and *M. hapla* Chitwood (Mh, northern RKN), are the cause of 95% of incidences of root-knot nematode parasitism on crops worldwide (Hussey and Janssen, 2002), and are able to invade over 2000 plant species, with species representing almost every plant family (Roberts, 1995; Trudgill and Blok, 2001). However, only Mi and Ma are of most concern to soybean growers in the southeastern USA (Boerma and Hussey, 1992; Rodríquez-Kábana et al., 1991). The wide dispersal of Mi (ie., the southern RKN), of particular concern in soybean fields in Georgia (Herman et al., 1991), has led to its distinction as possibly “the single most damaging crop pathogen in the world” (Trudgill and Blok, 2001). Ma (ie., the peanut RKN) is found within the same latitudinal distribution as Mi, but it is not as common as either Mi or Mj. As shown by its morphology, biochemistry, and cytological profile, Ma is also a diverse and unique RKN from Mi, Mh, and Mj (Cliff and Hirschmann, 1985; Eisenback and Triantophyllou, 1991; Hugall et al., 1994).

Since most southern breeding programs for soybean, tobacco (*Nicotiana tabacum* L.), and cotton (*Gossypium* L.) have traditionally focused on reducing damage from Mi, there is evidence that damage from Ma to Ma-susceptible crops such as soybean, corn, and tobacco is increasing (Fassuliotis, 1987; Ibrahim and Lewis, 1993). This is particularly the case in the southeastern states of Florida (Garcia and Rich, 1985), South Carolina (Fortnum et al., 1984), Alabama (Rodríguez-Kábana and Williams, 1981), and North Carolina (Schmitt and Barker, 1988). Of the two races of Ma that can infect soybean, race 1 is a persistent problem because its
major host, peanut, is grown on substantial hectareage in the southeastern USA and may be rotated with soybean (Rodríguez-Kábana et al., 1991). While race 2 of Ma does not parasitize peanut, it does appear to cause greater yield reduction on soybean than race 1 (Pedrosa et al., 1994) and includes several aggressive isolates (Carpenter and Lewis, 1991; Hiatt III et al., 1988; Ibrahim and Lewis, 1993). Race 2 is also more widely distributed, accounting for 84% of worldwide Ma populations (Riggs, 1991). Therefore, since many southern soybean cultivars do not possess adequate Ma resistance to either race (Kinloch et al., 1990), and Ma appears to be less sensitive to nematicides and more aggressive on soybean than Mi (Ibrahim and Lewis, 1993; Starr, 1989), studies focused on increasing the level of Ma resistance in soybean are important to improve the effectiveness of resistant cultivars, particularly in fields with high initial populations of Ma.

Novel sources of Ma resistance have been identified in the soybean germplasm collection. A comprehensive study of 2370 PIs identified PI 200538 and PI 230977 as highly resistant to both Ma (race 2) and Mj (Luzzi et al., 1987). The level of Ma resistance found in both PI sources was higher than that in the resistant check cultivar Jackson. While neither PI differed in gall index and Ma reproduction in the greenhouse, PI 200538 was less damaged by Ma in the field than was PI 230977 (Pedrosa et al., 1994). Mapping studies further evaluated F2:3 populations of PI 200538 x ‘CNS’ (a cultivar that is Ma-susceptible) for Ma resistance and identified a major resistance QTL on LG-F and a minor QTL on LG-E, which together accounted for 51% of the variation in gall number caused by Ma parasitism (Tamulonis et al., 1997a).

A more recent study established PI 594403, PI 594427C, and PI 594651L2 as potentially unique sources of resistance to Ma (Harris et al., 2003). Although these lines had more galls than PI 200538, both PI 594403 and PI 594427C had fewer or equal numbers of eggs,
respectively. While the genes conditioning the nematode resistance in PI 200538 and PI 230977 have been transferred into the respective breeding lines G93-9106 and G93-9223 (Luzzi et al., 1996; Luzzi et al., 1997), the resistance in PI 594403, PI 594427C, and PI 594651L2 has not yet been fully characterized. This study was therefore undertaken to evaluate these lines as potential sources of novel Ma resistance genes by evaluating the progenies of PI 200538 crossed to each of the three PIs for segregation of Ma resistance.

**Materials and Methods**

**General Design**

Information about the plant material evaluated in this experiment appear in the next section. All plants were grown in 20.6-cm individual Ray Leach Single-Cell Conetainers (Model LD-UV SC-10, Stuewe & Sons, Inc., Corvallis, OR). The individual conetainers (hereafter called cones) were arranged in every other column of Ray Leach 98-cone trays, with each tray consisting of seven columns and seven rows (or 49 cones tray$^{-1}$). The cones were filled with soil, which was Pacolet sandy loam and belongs to the fine, kaolinitic, thermic family of Typic Kanhapludults. This was supplemented with sand to a final weight weight$^{-1}$ composition of 73% sand, 16% silt, and 11% clay. The greenhouse was kept at a temperature of approximately 28º C, and 15 hours of light from 400-W Multivapor metal halide lamps (Westinghouse Electric Corp., Lamp Division, Bloomfield, NJ) was provided. The plants were watered manually three times daily, and were watered weekly with a solution supplemented with 6 mg N, 3 mg P, and 5 mg K.

A single plant was evaluated in each cone. In soybean, root gall numbers and egg density were found to have a positive correlation of 0.6 to 0.7 (Luzzi, 1986); therefore, this study interpreted the extent of galling as evidence of resistance or susceptibility. Since there is often a
weak association between the gall number of single F2 plants and their derived families (Luzzi et al., 1994; Luzzi et al., 1995a; 1995b), the results from each F2 screen were confirmed twice with replicated F2:3 families and, in some cases, F2:4 families.

_Evaluation of F2 Plants_

The PI 200538 x PI 594427C, PI 200538 x PI 594403, and PI 200538 x PI 230977 F2 populations were each evaluated in seven trays, with each cone planted with two seeds and thinned a week later to one plant per cone. The experiment for three of the four populations consisted of 343 plants, including 287 F2 plants and 56 checks [which included the parents, CNS (Ma-susceptible), and ‘Haskell’ (moderately Ma-resistant)]. Due to limited seed, only 164 F2 plants and 32 check plants were evaluated in the PI 200538 x PI 594651L2 population. For each experiment, additional plants of CNS, Haskell, ‘GaSoy-17’ (Ma-susceptible), ‘Bryan’ (moderately Ma-resistant), and ‘Perrin’ (moderately Ma-resistant) were planted in a separate tray to determine the optimal time to rate the plants in the experiment. Each plant was inoculated with approximately 5000 Ma race 2 eggs 7 d after planting. The eggs were obtained from the root systems of the susceptible eggplant (Solanum melongena L.) cultivar Black Beauty, and were collected with 0.5% NaOCl as described previously (Hussey and Janssen, 2002). The Ma inoculum was dispensed using a digital dispensing pump (Asteck Systems Corp., Farmingdale, NY).

Approximately 4 wk after inoculation, the plants were removed from the cones, and the roots were rinsed and examined for gall numbers. Twenty resistant F2 plants and 20 susceptible F2 plants from each population were transplanted into 30.5-cm pots to produce F2:3 seeds. The F2:3 seeds collected from the plants were used to confirm the results in two additional greenhouse experiments.
The subsequent F$_{2:3}$ lines from each population were evaluated in a randomized complete block design with two replications. The experimental unit consisted of seven cones of each line. For the PI 200538 x PI 594403 cross, four susceptible and nine resistant lines were tested, with two entries each of the parents, CNS, and Haskell within each replication. For the PI 200538 x PI 230977 population, 10 resistant and seven susceptible lines were evaluated, with one entry each of the parents, CNS, and Haskell within each replication. In the PI 200538 x PI 594427C test, 12 resistant and six susceptible lines were included, along with one entry each of the two parents and CNS. In the PI 200538 x PI 594651L2 cross, 10 resistant and eight susceptible lines were tested, including one entry of each of the parents and CNS in each replication. The F$_{2:3}$ plants from each experiment were inoculated with approximately 6000 Ma eggs 7 d after planting. For those F$_{2:3}$ lines that did not have adequate seed for two replications of seven plants each, seeds of ‘Benning-RR’ were substituted but not rated for galling. The experiment was terminated approximately 4 wk after inoculation, at which point the plants were removed from the cones, the roots were rinsed to remove soil debris, and the number of galls were determined.

Certain F$_{2:3}$ lines had to be increased to generate enough seeds to use in the second confirmation experiment. Between four and eight F$_{3}$ plants of each line were transplanted into larger pots, the seeds of which were bulked to obtain F$_{2:4}$ seeds.

In the second confirmation experiment, the lines were grown in a randomized complete block design and the experimental unit was one plant of each line in a cone. For the populations from PI 200538 x PI 594427C and PI 200538 x PI 594651L2, each block contained entries from three resistant families and two susceptible families from each cross, three entries of each of PI 200538, PI 594427C, and PI 594651L2, and one entry each of CNS and Haskell in a total of 15
replications. In the PI 200538 x PI 594403 and PI 200538 x PI 230977 crosses, four resistant and four susceptible entries from the former population and six resistant and six susceptible entries from the latter population were tested in 12 replications. One entry of each of the parents and CNS were included in each replication. The plants within each experiment were inoculated with 7500 Ma eggs 7 d after planting, and were evaluated for the number of galls on the roots approximately 4 wk later.

The galling data from the two confirmation experiments were analyzed with SAS (v.8, Cary, NC) using the PROC GLM procedure. Due to the heterogeneous error variance, the gall numbers were transformed with the square-root function. Single degree-of-freedom contrasts were performed between the lines from each cross that had been characterized in earlier generations as resistant or susceptible.

Results

Wide variation in galling was observed among the F2 individuals in PI 200538 x PI 594427C (Fig. 3-1). Fairly wide variation for galling was also found within the three other populations, although in each case it was not as extreme as that seen in the individuals from the PI 200538 x PI 594427C population. As shown in Figure 3-1, the highly galled plant had a small root system, comparable in size to CNS, while the roots of the plant with fewer galls were dense and healthy. This was a trend that was often observed in the experiment. The presence of both highly resistant plants and moderately galled plants within each population indicated that the two parents may contain unique Ma resistance genes that are segregating among their F2 progeny. Although the extent of the segregation varied among the progeny from the different crosses, it was large enough in each case to justify their inclusion in the confirmation experiments. Within
each population, the plants with the most extreme gall numbers were transplanted and their progeny were evaluated for their level of Ma resistance in the confirmation experiments.

The results from the lines in both confirmation experiments supported the initial results from the F2 individuals in all four of the crosses. The analysis of variance revealed highly significant differences (P<0.0001) between the lines in both confirmation experiments in the populations from PI 200538 x PI 594403, PI 200538 x PI 594427C, and PI 200538 x PI 594651L2, and significant differences (P<0.03) among the F2-derived progeny of PI 200538 x PI 230977 in both confirmation experiments. Single degree-of-freedom contrasts between the F2-derived lines classified as resistant or susceptible based on their initial F2 phenotype were statistically different (P<0.0001) in all populations, except for the PI 200538 x PI 230977 population (P>0.4 in both confirmation experiments, Table 3-1). Although the galling in the classes of resistant and susceptible lines could be significantly distinguished in two experiments, the extent of galling for the susceptible F2:3 lines did not ever equal or exceed that observed in Ma-susceptible CNS, which consistently averaged at least 100 galls. The one exception was a line from the PI 200538 x PI 594403 cross. This suggests that the majority of lines, although varying in their levels of resistance, had one or more resistance genes in common.

The ranges of galling in the resistant and susceptible F2-derived lines from PI 200538 x PI 594427C and PI 200538 x PI 594651L2 had little or no overlap in the confirmation experiments (Table 3-1). Although the contrast of the resistant and susceptible lines from the PI 200538 x PI 594403 cross also showed significant differences (P<0.0001), their ranges of galling did overlap in the first experiment, but not in the second (Table 3-1). This indicates that while the F2 phenotype of the PI 200538 x PI 594403 plants was overall predictive of the F2-derived galling response, there may have been environmental variation in the F2 experiment that
obscured the results in confirmation experiments (eg., F2 plants were misclassified). Since fewer lines were tested in the second confirmation experiment, some of the misclassified lines may have unintentionally been excluded, resulting in nonoverlapping ranges in that experiment. With the exception of the PI 200538 x PI 230977 experiment, the highly galled F2 individuals in three of the four crosses had F2:3 and F2:4 progeny that had more severe galling than did the F2:3 and F2:4 progeny from the F2 individuals with low galling.

Single lines within each of the gall classes showed significant differences in galling from each other and from the parents. Within the PI 594427C x CNS population, the line 460-3 showed less than half the number of galls than was observed on either parent (Table 3-2). This effect was consistent in the two confirmation experiments, and indicates that this line may contain unique resistance alleles at independent loci from each parent. Line 465-1, which also belonged to this population, had approximately twice as many galls on its roots than was observed on either parent in the first confirmation experiment (Table 3-2), indicating that this line, in contrast to line 460-3, did not inherit as many resistance alleles from each parent. This line, along with other high-galling lines that were tested in the first confirmation experiment, did not produce enough seeds to be tested in the second confirmation experiment. For this reason, none of the lines evaluated in the second confirmation experiment of this population demonstrated the extent of galling that was observed in 465-1.

Single lines with variable Ma reactions were also observed in the other populations. Line 477-1 from the PI 200538 x PI 594651L2 population consistently contained 10 fewer galls on its root system compared to PI 594651L2 in both confirmation experiments, and to PI 200538 in the first confirmation experiment. Line 469-1, in contrast, contained from 8 to 22 more galls compared to either parent (Table 3-2). In the PI 200538 x PI 230977 population, in which
significant differences between the resistant and susceptible classes were not observed, line 401-3 had fewer galls than either parent, line 397-1 had fewer galls than PI 200538, and both lines had approximately half the number of galls than was observed in lines 402-1 and 393-1 (Table 3-2). The latter lines also averaged more than 20 galls than either parent. In contrast, no lines were recovered in the PI 200538 x PI 594403 population that had significantly fewer galls than PI 200538; however, several lines, such as 385-1 and 390-1, were recovered that had significantly more galls than either parent (Table 3-2).

Discussion

The procedure of screening progenies of Ma-resistant x Ma-resistant crosses for potential segregation of resistance was previously used for evaluating Jackson x PI 200538 and PI 200538 x PI 230977 populations (Luzzi et al., 1995a; 1995b). In these studies, F2 individuals and F2:3 lines from each population were screened for resistance, with checks including both susceptible lines and lines with moderate to high levels of resistance. The presence of susceptible F2 plants, confirmed by two evaluations of derived F3 families, suggested that the parents of both crosses contained unique resistance genes.

The data from the two independent experiments in the PI 200538 x PI 594427C, PI 200538 x PI 594651L2, and PI 200538 x PI 594403 populations in this study also suggests the presence of unique resistance genes. In both confirmation experiments, the resistant lines of each cross in these populations showed significantly less galling than the susceptible lines, although in the PI 200538 x PI 594403 cross the presence of overlapping ranges of the resistant and susceptible groups in the first confirmation test suggested a misclassification of some of the original F2 plants. Of the lines that were used in both confirmation experiments, most had similar galling results in both tests, again demonstrating the consistency of the Ma response. In
contrast, the classes of resistant and susceptible lines in the PI 200538 x PI 230977 population were not significantly different. These two PIs contain major QTLs at the same relative position on LG-F (Tamulonis et al., 1997a; Tamulonis et al., 1997b). Although the QTLs were mapped for resistance to two different species of RKN (PI 200538 for Ma resistance and PI 230977 for Mj resistance), both PIs are resistant to both Ma and Mj (Luzzi, 1986). It is not illogical to assume that the two PIs do in fact contain Ma-resistance alleles at the same QTL, and that they possess the same allele that conditions resistance to both Ma and Mj. If that is the case, then PI 200538 and PI 230977 differ only in the location of minor QTLs, with PI 200538 containing a minor QTL on LG-E and PI 230977 containing a minor QTL on LG-D1b (Tamulonis et al., 1997a; Tamulonis et al., 1997b). It is possible that this subtle variation, caused by the segregation of alleles with minor effects in the different lines, was not distinguished in the tests between the classes, but could be resolved on a single-line basis.

The correspondence of Ma response of the F2 plants compared to their F2:3 progeny shows that Ma resistance is heritable on a single-plant basis. Even stronger support for this claim is shown by the recovery of single lines that showed transgressive segregation from the parents. In the population PI 200538 x PI 594427C, the line 460-3 consistently had half the number of galls than were observed in either parent, and had even fewer galls compared to lines 463-1 and 465-1. These latter lines contained more galling than either parent in at least one of their confirmation experiments. The difference, however, was much greater in the first experiment than in the second. Due to decreased viability and low seed production in the highly galled lines in this population, it was difficult to obtain enough seeds of the most severely galled lines for use in both confirmation experiments.
Consistent delineations between highly galled and highly resistant lines were also observed in the PI 200538 x PI 594651L2 and PI 200538 x PI 230977 populations. In the former population, this effect was even stronger since the same lines in both confirmation experiments, 469-1 and 477-1, showed consistent difference in Ma response compared to each other and to the parents. The results in the latter population were particularly notable since significant differences were not observed between resistant and susceptible lines when considered as different classes. This result, taken on its own, would indicate that the resistance in the two PIs is completely allelic. However, the presence of single lines within the population that do differ in resistance and susceptibility, suggest that the two PIs do contain unique resistance alleles at different loci.

Overall, these differences indicate that PI 230977, PI 594427C, PI 594651L2, and possibly PI 594403, each contain Ma-resistance gene(s) that are unique from those in PI 200538. Since allelic variation at the same locus cannot be easily distinguished from alleles at independent loci (ie. different genomic positions) in this study, it is also possible that these PIs contain different alleles at the same major locus on LG-F that could also provide a unique or more effective source of Ma resistance. While additional tests are required, the magnitude of the differences between the PIs in each of the populations can give an indication of the magnitude of effect of unique resistance genes indicated by this study. Single lines that were highly resistant in the PI 200538 x PI 230977 and PI 200538 x PI 594651L2 populations did not differ in galling from the parents as much or as consistently as the line in the PI 200538 x PI 594427C population. This may indicate that the unique resistance gene(s) in PI 594427C may have a greater effect on Ma resistance than that in PI 230977 or PI 594651L2.
PI 594403 had initially been reported as a highly Ma-resistant accession (Harris et al., 2003). However, in this experiment a high level of Ma galling in PI 594403 was observed. Additionally, it was not possible to recover lines within the PI 200538 x PI 594403 population with fewer galls than PI 200538, suggesting that PI 594403 does not contain a major-effect resistance gene. High galling had also been observed on PI 594403 in earlier greenhouse assays, and it is possible that this seed source is heterogeneous for Ma resistance. In agreement with this, the progeny from this cross also revealed more variation in galling than in the other crosses, to the extent that the ranges of galling between the resistant and susceptible lines overlapped in one confirmation experiment. PI 594403 may still contain unique resistance genes, but it may be necessary to purify the seed source before further characterization is possible.

The novel resistance genes that are potentially present in PI 230977, PI 594427C, PI 594651L2, and PI 594403 can complement that provided by the resistance allele introgressed from PI 200538, resulting in a higher level of Ma resistance than is currently available. In order to best utilize these resistance genes in a soybean breeding program, it is important to understand their genomic locations and potential intra-locus effects.

Acknowledgements

The authors gratefully acknowledge greenhouse assistance from Charles Lester, Matthew Ephgrave, Earl Baxter, Matt Wood, Tasaday Lynch, and Erika Swinson. This work was supported by funding from the United Soybean Board, the Georgia Agricultural Experiment Stations, and an NSF pre-doctoral fellowship to J.L. Yates.

References


Figure 3-1. Segregation of *M. arenaria* galling among F$_2$ plants from the resistant x resistant cross PI 594427C x PI 200538. CNS is presented as a Ma-susceptible check.
Table 3-1. Mean and range in *M. arenaria* (Ma) gall number in two confirmation experiments for F$_{2.3}$ lines derived from least galled and most galled F$_2$ plants of four Ma-resistant x Ma-resistant crosses.

<table>
<thead>
<tr>
<th>Cross</th>
<th>F$_2$ gall class / parent</th>
<th>Confirmation Test 1</th>
<th>Confirmation Test 2</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Entries</td>
<td>Mean</td>
</tr>
<tr>
<td>PI 200538 x PI</td>
<td>Least galls</td>
<td>12</td>
<td>$22^{b\dagger}$</td>
</tr>
<tr>
<td>594427C</td>
<td>Most galls</td>
<td>6</td>
<td>$58^{a}$</td>
</tr>
<tr>
<td></td>
<td>PI 200538</td>
<td>1</td>
<td>33</td>
</tr>
<tr>
<td></td>
<td>PI 594427C</td>
<td>1</td>
<td>44</td>
</tr>
<tr>
<td>CNS</td>
<td></td>
<td>1</td>
<td>119</td>
</tr>
<tr>
<td>PI 200538 x PI</td>
<td>Least galls</td>
<td>10</td>
<td>$26^{b}$</td>
</tr>
<tr>
<td>594651L2</td>
<td>Most galls</td>
<td>8</td>
<td>$43^{a}$</td>
</tr>
<tr>
<td></td>
<td>PI 200538</td>
<td>1</td>
<td>30</td>
</tr>
<tr>
<td></td>
<td>PI 594651L2</td>
<td>1</td>
<td>32</td>
</tr>
<tr>
<td>CNS</td>
<td></td>
<td>1</td>
<td>185</td>
</tr>
<tr>
<td>PI 200538 x PI</td>
<td>Least galls</td>
<td>9</td>
<td>$46^{b}$</td>
</tr>
<tr>
<td>594403</td>
<td>Most galls</td>
<td>4</td>
<td>$59^{a}$</td>
</tr>
<tr>
<td></td>
<td>PI 200538</td>
<td>2</td>
<td>27</td>
</tr>
<tr>
<td></td>
<td>PI 594403</td>
<td>2</td>
<td>61</td>
</tr>
<tr>
<td>CNS</td>
<td></td>
<td>2</td>
<td>107</td>
</tr>
<tr>
<td></td>
<td>Least galls</td>
<td>Most galls</td>
<td>PI 200538</td>
</tr>
<tr>
<td>----------------</td>
<td>-------------</td>
<td>------------</td>
<td>------------</td>
</tr>
<tr>
<td><strong>PI 200538 x PI 230977</strong></td>
<td>10</td>
<td>7</td>
<td>1</td>
</tr>
<tr>
<td>Least galls</td>
<td>50&lt;sup&gt;a&lt;/sup&gt;</td>
<td>51&lt;sup&gt;a&lt;/sup&gt;</td>
<td>46</td>
</tr>
<tr>
<td>Most galls</td>
<td>28-65</td>
<td>39-73</td>
<td>1</td>
</tr>
<tr>
<td>PI 200538</td>
<td>6</td>
<td>6</td>
<td>1</td>
</tr>
<tr>
<td>PI 230977</td>
<td>37&lt;sup&gt;a&lt;/sup&gt;</td>
<td>38&lt;sup&gt;a&lt;/sup&gt;</td>
<td>33</td>
</tr>
<tr>
<td>CNS</td>
<td>23-54</td>
<td>31-48</td>
<td></td>
</tr>
</tbody>
</table>

† Values followed by the same letter do not significantly differ at the P=0.05 level.
Table 3-2. Mean *M. arenaria* gall number of lines from four Ma-resistant x Ma-resistant crosses in two different confirmation experiments.

<table>
<thead>
<tr>
<th>Cross</th>
<th>Line</th>
<th>Mean</th>
<th>Line</th>
<th>Mean</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Confirmation Test 1</strong></td>
<td></td>
<td>galls plant(^{-1}) ± SE</td>
<td><strong>Confirmation Test 2</strong></td>
<td></td>
</tr>
<tr>
<td>PI 200538 x PI</td>
<td>460-3</td>
<td>13 ± 2</td>
<td>460-3</td>
<td>13 ± 2</td>
</tr>
<tr>
<td>594427C</td>
<td>465-1</td>
<td>76 ± 4</td>
<td>463-1</td>
<td>38 ± 2</td>
</tr>
<tr>
<td>PI 200538</td>
<td>33 ± 3</td>
<td></td>
<td>PI 200538</td>
<td>28 ± 2</td>
</tr>
<tr>
<td>PI 594427C</td>
<td>44 ± 4</td>
<td></td>
<td>PI 594427C</td>
<td>40 ± 2</td>
</tr>
<tr>
<td>CNS</td>
<td>119 ± 15</td>
<td></td>
<td>CNS</td>
<td>134 ± 12</td>
</tr>
<tr>
<td>PI 200538 x PI</td>
<td>477-1</td>
<td>19 ± 2</td>
<td>477-1</td>
<td>27 ± 3</td>
</tr>
<tr>
<td>594651L2</td>
<td>469-1</td>
<td>52 ± 3</td>
<td>469-1</td>
<td>46 ± 5</td>
</tr>
<tr>
<td>PI 200538</td>
<td>30 ± 5</td>
<td></td>
<td>PI 200538</td>
<td>28 ± 2</td>
</tr>
<tr>
<td>PI 594651L2</td>
<td>32 ± 4</td>
<td></td>
<td>PI 594651L2</td>
<td>38 ± 3</td>
</tr>
<tr>
<td>CNS</td>
<td>185 ± 7</td>
<td></td>
<td>CNS</td>
<td>134 ± 12</td>
</tr>
<tr>
<td>PI 200538 x PI</td>
<td>387-3</td>
<td>26 ± 2</td>
<td>387-3</td>
<td>30 ± 4</td>
</tr>
<tr>
<td>594403</td>
<td>385-1</td>
<td>91 ± 6</td>
<td>390-1</td>
<td>139 ± 10</td>
</tr>
<tr>
<td>PI 200538</td>
<td>27 ± 2</td>
<td></td>
<td>PI 200538</td>
<td>33 ± 5</td>
</tr>
<tr>
<td>PI 594403</td>
<td>61 ± 3</td>
<td></td>
<td>PI 594403</td>
<td>98 ± 9</td>
</tr>
<tr>
<td>CNS</td>
<td>107 ± 5</td>
<td></td>
<td>CNS</td>
<td>134 ± 12</td>
</tr>
<tr>
<td>Combination</td>
<td>401-3</td>
<td>39 ± 4</td>
<td>397-1</td>
<td>23 ± 2</td>
</tr>
<tr>
<td>-------------------</td>
<td>-------</td>
<td>--------</td>
<td>-------</td>
<td>--------</td>
</tr>
<tr>
<td>230977 x PI 200538</td>
<td>402-1</td>
<td>73 ± 10</td>
<td>393-1</td>
<td>54 ± 16</td>
</tr>
<tr>
<td>PI 200538</td>
<td>46 ± 3</td>
<td>PI 200538</td>
<td>33 ± 5</td>
<td></td>
</tr>
<tr>
<td>PI 230977</td>
<td>48 ± 4</td>
<td>PI 230977</td>
<td>27 ± 4</td>
<td></td>
</tr>
<tr>
<td>CNS</td>
<td>123 ± 10</td>
<td>CNS</td>
<td>134 ± 12</td>
<td></td>
</tr>
</tbody>
</table>
CHAPTER 4

QUANTITATIVE TRAIT LOCI ASSOCIATED WITH SOYBEAN RESISTANCE TO PEANUT ROOT-KNOT NEMATODES IN PI 594403, PI 594427C, AND PI 594651L2

Abstract

Peanut root-knot nematode (*Meloidogyne arenaria* (Neal.) Chitwood or Ma) is one of the four most economically important RKN species (*Meloidogyne* spp.), and is increasing in population density in the southeastern USA. PI 200538, a soybean (*Glycine max* (L.) Merr.) plant introduction (PI), was previously found to contain a major Ma resistance quantitative trait locus (QTL) on LG-F and a minor QTL on LG-E. This study determined whether three recently characterized PIs also contained these QTLs, and, additionally, whether they might also contain novel resistance genes. In two of the three soybean PIs, PI 594427C and PI 594651L2, a QTL in the interval between L20a and Satt510 on LG-F was discovered that accounted for as much as 41% and 17% of the Ma galling response in two different populations (F2 and F2:3) created from each of the PIs crossed to the Ma-susceptible cultivar CNS. In addition, a QTL on LG-E that accounted for up to 6% of the variation was mapped in the F2 population of PI 594651L2 x CNS, and was confirmed in the F2:3 population of the same cross. A QTL on LG-E, which explained 7% of the phenotypic variation, was also mapped to approximately the same position in the F2 population of PI 594427C x CNS, but this QTL could only be detected when L20a was fixed for homozygosity, and was not detected in the F2:3 population of the same cross. Although QTLs were discovered in the F2 population of PI 594403 x CNS, they did not map to LG-F or LG-E, and none of them were confirmed in the F2:3 generation; therefore, this PI may be heterogeneous for Ma resistance. Both the QTL on LG-F and the QTL on LG-E reported in this study mapped to approximately the same position as the QTLs initially detected in PI 200538. These results confirm the importance of these resistance alleles on LG-F and LG-E, and establish the necessity for creating more high-throughput markers to allow soybean breeders to more easily introgress this source of Ma resistance into their cultivars.
Introduction

The RKNs belong to the genus *Meloidogyne*, which contains over 54 described species (Trudgill and Blok, 2001). The four most common RKN species, which include *M. incognita* (Kofoid and White) Chitwood (Mi), *M. arenaria* (Neal) Chitwood (Ma), *M. hapla* Chitwood (Mh), and *M. javanica* (Treub) Chitwood (Mj) are the cause of 95% of incidences of root-knot nematode parasitism on crops worldwide (Hussey and Janssen, 2002), and are able to invade over 2000 plant species, with species representing almost every plant family (Roberts, 1995; Trudgill and Blok, 2001). RKNs are responsible for more damage to the global food supply than any other group of plant-parasitic nematodes (Bernard, 1989), and caused 13% yield loss to the soybean crop in Florida in one year (Garcia and Rich, 1985). The wide host range of RKNs means that any crop grown in rotation with soybean is probably susceptible as well, including most of the major crops grown in the southeastern USA (Luzzi, 1986). Mi and Ma are of most concern to soybean growers in the southeastern USA (Boerma and Hussey, 1992; Rodríquez-Kábana et al., 1991). Ma is found within the same latitudinal distribution as Mi, but is not as common as Mi in soybean fields in Georgia (Herman et al., 1990). However, its morphology, biochemistry, and cytological profile are more variable and atypical than Mi (Cliff and Hirschmann, 1985; Eisenback and Triantophyllou, 1991). This variability has resulted in great variability in the hosts’ response to this parasite (Kinloch et al., 1987), which complicates the development of Ma resistant cultivars.

More effort has been focused on obtaining high levels of resistance to Mi, not just in soybean, but also in tobacco (*Nicotiana tabacum* L.) and cotton (*Gossypium hirsutum* L.). The predominant culture of Mi-resistant cultivars may have increased non-target populations of Ma (Fassuliotis, 1987; Garcia and Rich, 1985; Ibrahim and Lewis, 1993). Beginning in the 1980s,
more studies begin to report increasing Ma populations in fields in the southeastern USA (Fortnum et al., 1984; Garcia and Rich, 1985; Rodríguez-Kábana and Williams, 1981; Schmitt and Barker, 1988). Few soybean cultivars adapted to this region have adequate Ma resistance to avoid yield losses (Garcia and Rich, 1985; Kinloch et al., 1990). Ma races and isolates of races can also vary widely in their aggressiveness and ability to reduce yield in soybean (Carpenter and Lewis, 1991; Hiatt III et al., 1988; Ibrahim and Lewis, 1993; Pedrosa et al., 1994). The identification of novel Ma resistance alleles is thus necessary to protect the soybean crop from even the most aggressive Ma isolates.

However, the narrow genetic base of elite soybean cultivars precludes significant variation for many traits, including RKN resistance. One statistic estimates that almost 80% of the diversity in elite U.S. soybean cultivars was contributed by fewer than 10 PIs (Delannay et al., 1983). Yet multiple resistance genes are needed in cultivars to provide more durable resistance, particularly since soybean fields in Georgia may be infested with different species of plant parasitic nematodes (Niblack et al., 1986). There are several sources of Ma resistance in the soybean germplasm. A comprehensive study of 2370 PIs identified PI 200538 and PI 230977 as highly resistant to both Ma (race 2) and Mj (Luzzi et al., 1987). PI 200538 was used in quantitative trait locus (QTL) mapping studies for Ma resistance. PI 230977, in contrast, has been the focus of mapping studies for resistance to Mj. These studies established that both PIs had major resistance QTLs in similar positions on LG-F, but had minor QTLs on different linkage groups (Tamulonis et al., 1997a; 1997b). The major QTL on LG-F and the minor QTL on LG-E in PI 200538 accounted for 51% of the variation in gall number caused by Ma in the F_{2:3} population, while the major QTL on LG-F and the minor QTL on LG-D1b in PI 230977 explained 54% of the variation in gall number caused by Mj.
Another more recent study established PI 594403, PI 594427C, and PI 594651L2 as promising sources of resistance to Ma (Harris et al., 2003). Although these new PIs had more galls than PI 200538, both PI 594403 and PI 594427C had fewer or equal numbers of eggs, respectively. While the genes conditioning the nematode resistance in PI 200538 and PI 230977 have been transferred into the respective breeding lines G93-9106 and G93-9223 (Luzzi et al., 1996; Luzzi et al., 1997), the resistance in PI 594403, PI 594427C, and PI 594651L2 has not yet been fully exploited. The segregation results in replicated experiments on progenies from several resistant x resistant crosses of these PIs to PI 200538 (Yates, 2006) suggested that these PIs contained unique resistance alleles from those in PI 200538. These unique alleles could map to the same loci on LG-F and LG-E, or could map to entirely new loci. Therefore, F2 mapping populations were developed from each of these PIs crossed to CNS as the common susceptible parent. The F2 mapping results were confirmed with F3 families that were derived from the same crosses, but were developed from different F2 plants.

Materials and Methods

F2 Mapping Populations

In 2003, crosses were made between each of PI 594403, PI 594427C, and PI 594651L2 (all of which are highly Ma-resistant) to the susceptible parent CNS. Seed chips were used to confirm F1 seeds to be the result of cross-pollination with the use of SSR (simple sequence repeat) markers. F2 seeds were harvested from each F1 plant in the late spring of 2004. F2 seeds were planted in equal amounts from each F1 family, with a total of 188 F2 plants analyzed for each of the three crosses.

All plants were grown in 20.6-cm individual Ray Leach Single-Cell Conetainers (Model LD-UV SC-10, Stuewe & Sons, Inc., Corvallis, OR). The individual conetainers (hereafter
called cones) were arranged in every other column of Ray Leach 98-cone trays, with each tray consisting of seven columns and seven rows (or 49 cones tray⁻¹). The cones were filled with soil, which was Pacolet sandy loam and belongs to the fine, kaolinitic, thermic family of Typic Kanhapludults. This was supplemented with sand to a final weight composition of 73% sand, 16% silt, and 11% clay. A single plant was evaluated in each cone. The plants were kept in a greenhouse maintained at approximately 28°C and were watered manually three times daily, which was supplemented with 6 mg N, 3 mg P, and 5 mg K on a weekly basis. Fifteen hours of light were provided by 400-W metal halide lamps (Westinghouse Electric Corp., Lamp Division, Bloomfield, NJ) suspended 1.4 m above the greenhouse benches.

The experiment for each population consisted of six trays, with a little over five trays containing F₂ plants and four checks per tray (two of each parent in each tray). The remaining trays consisted of additional checks of CNS, PI 200538, ‘Perrin’ (Ma-resistant), ‘Bryan’ (Ma-resistant), ‘GaSoy17’ (Ma-susceptible), and ‘Haskell’ (Ma-resistant). The F₂ plants were treated with two separate inoculations of 3000 Ma eggs each, starting approximately 7 d after planting, with each inoculation separated by 2 or 3 d. Eggs were collected with 0.5% NaOCl from the root systems of the susceptible eggplant (Solanum melongena L.) cultivar Black Beauty as described previously (Hussey and Janssen, 2002), and were dispensed using a digital dispensing pump (Asteck Systems Corp., Farmingdale, NY). Approximately 30 d after inoculation, the plant tops were removed, the roots were rinsed free of soil debris, and the galls per individual root system were counted. Small trifoliolate leaves were collected from F₂ individuals on several different occasions to extract DNA for SSR marker analysis.

F₂:3 Mapping Populations
F$_{2;3}$ populations were created for the same crosses but from different F$_2$ plants than were used in the F$_2$ mapping populations. In January 2005, approximately 300 F$_2$ seeds from each population (PI 594403 x CNS, PI 594427C x CNS, and PI 594651L2 x CNS) were sent to the USDA winter nursery in Puerto Rico. Upon reaching maturity, the F$_3$ seeds from 168 to 173 F$_2$ plants from each population were single-plant thresher. The seeds were used to plant two replications of the 168 to 173 F$_2;3$ lines in 1-L Styrofoam cups filled with standard potting soil. The experimental design was a randomized complete block. Six seeds of each F$_2;3$ line were initially planted in each cup, and the number of seedlings was later thinned to four per cup. The checks included the two parents and Haskell (which is moderately Ma-resistant). In total, there were approximately 200 cups per replication, 400 cups per experiment, and 1200 cups total. The F$_3$ families from each of PI 594403 x CNS, PI 594427C x CNS, and PI 594651L2 x CNS were staggered in planting dates by at least 1 wk. Approximately 1 wk after germination, the plants in each cup were inoculated with a total of 10,000 Ma eggs that were collected as previously described. During the experiment, trifoliolate leaves from each of the eight plants representing a family were collected to use in DNA extraction for the SSR marker analyses. Approximately 30 d after inoculation, the roots of the plants were rinsed and rated for the extent of galling on the roots.

Rather than counting the number of galls on each of the four plants in a cup, a rating scale from 1 to 5 (1=very few galls and 5=very many galls) was developed using cultivars and breeding lines with known levels of resistance. A week prior to planting the confirmation experiment, an experiment was planted in order to establish this scale. The cultivars and breeding lines included G93-9106 (highly Ma-resistant), ‘Haskell-RR’ (moderately Ma-resistant), ‘Boggs-RR’ (moderately Ma-resistant), ‘Benning-RR’ (moderately Ma-resistant),
‘Prichard-RR’ (Ma-susceptible) and CNS (Ma-susceptible). Ten cups containing four seeds each were planted for each of the lines. These lines were inoculated and their roots harvested as described above. Roots with none or very few galls were chosen to represent a rating of “1”, while roots with extremely heavy galling represented a rating of “5”. Roots that had levels of galling varying between those extremes were chosen to represent “2”, “3”, and “4” (Fig. 4-1). Photographs were taken of all categories to assist in rating the plants from the confirmation experiments.

**DNA Extraction and SSR Marker Analyses**

Trifoliolate leaves were collected from each of the plants in the F2 and F2:3 on several occasions. The leaves were placed in coin envelopes and freeze-dried for at least 48 h. One leaf from each plant in the F2 populations was placed in an individual well of a 96-well deep-well plate, along with a zinc-plated 4.5-mm BB (Daisy Outdoor Products, Rogers, AR). In the F2:3 populations, eight leaves were transferred into individual wells of a 48-well deep-well plate that already contained BBs in each well. Each plate was placed in a custom-made leaf grinder and shaken (exact RPM unknown) for approximately 1 min to pulverize the tissue. DNA extraction then proceeded as described in Keim et al. (1988). The DNA pellet was resuspended in TE buffer (10 mM Tris-HCl, 1 mM EDTA pH 8), and adjusted to a concentration of approximately 50 ng µL⁻¹ for use in PCR (polymerase chain reaction). PCR conditions followed guidelines recommended for SSR markers that were established for soybean by Diwan and Cregan (1997). The SSRs had a blue (6-FAM), yellow (NED), or green (HEX) fluorescent tag label (PE-ABI, Foster City, CA). PCR cycling conditions were the same as those used in Diwan and Cregan (1997) with a PE-ABI 9700 thermocycler (PE-ABI, Foster City, CA) with 384-well heating blocks or a MJ Research Tetrad thermocycler (Bio-Rad, Inc. Watertown, MA). The PCR
products were resolved on a DNA sequencer (ABI PRISM 377, PE-ABI, Foster City, CA). GeneScan® 3.1.2 software (PE-ABI, Foster City, CA) was used to determine the size of the products using the ROX 500 ladder (PE-ABI, Foster City, CA) as the size standard.

Other PCR-based markers used in this study were resolved either on a 4% agarose gel and visualized with ethidium bromide, or by use of the single-strand conformation polymorphism (SSCP) technique and visualized with silver staining. The following paragraph pertains in particular to the PCR-based markers OPN11, L20a, and 83I9. L20a and OPN11 are each described in previous reports (Hayes et al., 2000; Zheng et al., 2003), although the primer sequences used to amplify the L20a segment are not described in the paper. The L20a primer sequences used in this study are F-5’-TGTGAGGCATCAGAAGGGG-3’ and R-5’-GCCACCTATGCCACAATCC-3’. The PCR conditions for OPN11 are described elsewhere (Zheng et al., 2003), while for L20a the PCR mix was a 25-µL volume consisting of 1X buffer (10 mM Tris-HCl pH 9.0, 50 mM KCl, 0.1% Triton®-X-100), 2.5 mM MgCl₂, 0.2 mM dNTPs, 1 unit Taq DNA polymerase (M1668, Promega Corp., Madison, WI), 0.4 µM of each primer, and approximately 50 ng µL⁻¹ DNA template. The PCR cycling conditions consisted of 30 cycles of a denaturation step at 94°C for 30 s, 54°C for 30 s, and 72°C for 1 min. Approximately 3 µL of the PCR products amplified with OPN11 and L20a were combined with 9 µL of a loading dye solution (950 µL deionized formamide, 40 µL of a 1% solution of each of bromophenol blue and xylene cyanol, and 10 µL 1 M NaOH) and denatured for 2 min at 95°C. The samples were loaded on a 0.5X MDE™ (Cambrex Bio Science, Rockland, Inc.) gel (71.5 mL H₂O, 27.5 mL 2X MDE, 11 mL 6X TBE, 800 µL 10% APS, 160 µL TEMED) on a vertical MegaGel polyacrylamide gel apparatus (C.B.S. Scientific Company, Solana Beach, CA) and electrophoresed at 100 V (1 W gel⁻¹) for approximately 40 hours. At the end of the run, the PCR
products were detected by silver staining, first by agitating in fixer (210 mL 95% ethanol and 10 mL glacial acetic acid brought to a 1-L volume) for 10 min, followed by 10 min in silver nitrate solution (2% silver nitrate in fixer), two quick washes with deionized H₂O, and treatment with developer (75 mL 10 N NaOH, 1 mL 37% formaldehyde brought to a 1-L volume) until the PCR fragments were at the desired intensity. The gel was then rinsed once with fixer to stop the reaction and was photographed.

The marker 83I9 is an SSR developed from a bacterial artificial chromosome (BAC) that was mapped to LG-F as part of an effort to sequence this part of the genome. The SSR was kindly provided by Dr. Saghai Maroof (Virginia Polytechnic University, Blacksberg, VA). PCR conditions consisted of a 10-µL volume of 1X buffer (10 mM Tris-HCl pH 9.0, 50 mM KCl, 0.1% Triton®-X-100), 1.5 mM MgCl₂, 0.16 mM dNTPs, 0.5 units Taq DNA polymerase (M1668, Promega Corp., Madison, WI), 0.2 µM of each primer, and approximately 50 ng µL⁻¹ DNA template. The PCR products were electrophoresed on a 4% low melting point agarose gel and visualized with ethidium bromide staining.

Statistical analysis

The gall scores for lines within replications were examined for consistent responses across the two replications. If the gall ratings between replications differed by a gall score of more than two, then that entry was removed from further analysis. The gall scores from remaining lines were then averaged across replications. PROC GLM in SAS (v.8, Cary, NC) was used to generate least-squared (LS) means of the galling scores in the F₂:₃ populations across the replications and to generate estimates for families that were not represented in both replications. The genotypic and phenotypic data were analyzed for the presence of potential QTLs using MapManager QTX (Manly et al., 2001). The maternal genotype was designated a
“1”, the paternal genotype a “3”, and the heterozygote a “2”. Linkage groups were constructed based on a significance level of P<0.05. Both marker regression and interval mapping were used to search for associations. In the case of the latter, the LOD (logarithm of odds) score that defined the level of significance was determined through a permutation using 5000 shufflings of the phenotypic and genotypic data.

Broad-sense heritability estimates was calculated as:

\[
\frac{\sigma_L^2}{(\sigma^2_{\varepsilon}/2) + \sigma_L^2}
\]

using variance component estimates generated through SAS (v. 8, Cary, NC), where \(\sigma^2_{\varepsilon}\) is the experimental error and \(\sigma_L^2\) is the F2:3 line variance component.

**Results and Discussion**

*Phenotypic Distribution*

As shown by the phenotypic distribution of the F2 populations of PI 594403 x CNS, PI 594427C x CNS, and PI 594651L2 x CNS, the progeny and parents exhibited significant variation in galling (Fig. 4-2). In each population, CNS averaged more galls than the Ma-resistant parents PI 594403, PI 594427C, and PI 594651L2. Transgressive segregation of progeny more susceptible than CNS was observed in all cases, but was not as large in the population PI 594403 x CNS (Fig. 4-2a). The progeny of PI 594403 x CNS, as well as those of PI 594651L2 x CNS, were both skewed towards resistance (Fig. 4-2a and 4-2c), while the population PI 594427C x CNS appeared to have a normal distribution of gall numbers (Fig. 4-2b).

The particular trends observed in the F2 phenotypic distributions of the three populations were similar to those observed in the F2:3 populations (Fig. 4-3). This suggests that the use of the
scale to phenotype the F_{2:3} populations, rather than counting the galls on each root system, was an effective system to categorize the extent of galling in each of the families. The F_{2:3} lines in all populations showed significant variation in galling, and the susceptible parent CNS always had a more severe gall rating than the resistant parent (Fig. 4-3). As in the F_{2} populations, the F_{2:3} populations of both PI 594403 x CNS and PI 594651L2 x CNS were skewed towards lower ratings (Fig. 4-3a and 4-3c), while that of PI 594427C x CNS showed evidence of a normal distribution of gall ratings (Fig. 4-3b). Transgressive segregation for F_{2:3} lines being more susceptible than CNS was observed in all populations (Fig. 4-3b). Broad sense heritability values based on the mean of two replications (mean of 8 plants) for Ma resistance in the F_{2:3} populations of PI 594403 x CNS, PI 594427C x CNS, and PI 594651L2 x CNS were 0.47, 0.69, and 0.63, respectively.

**Identified QTLs**

Four hundred and sixty SSR markers were used to detect polymorphisms between CNS and PI 594403, PI 594427C, or PI 594651L2. The three populations in this mapping study had relatively high levels of polymorphisms; PI 594403 and CNS showed the most polymorphisms (49%) with SSR markers, while PI 594427C and CNS and PI 594651L2 x CNS had similar levels of polymorphisms (approximately 45% in both cases). A subset of the polymorphic markers were tested on the F_{2} progeny of the three populations, and were selected to provide an approximate even coverage over all 20 linkage groups based on the SSR marker order in the soybean genetic map found in Song et al. (2004). Data were collected for a total of 196 markers in the PI 594403 x CNS population, 180 markers in the PI 594427C x CNS population, and 153 markers in the PI 594651L2 x CNS population. The distance between mapped markers on each of the linkage groups was approximately 10.6 cM in the populations, and a range of 4 to 17
markers were mapped on each of the linkage groups in the populations. More markers were tested and mapped from linkage groups known from mapping results from PI 200538 to be associated with Ma resistance, such as LG-F and LG-E.

Relatively large-effect QTLs were found in similar positions on LG-F in the populations of PI 594427C x CNS and PI 594651L2 x CNS (Fig. 4-4). The QTL in the population of PI 594427C x CNS was most closely associated with the 0.9-cM interval between 83I9 and L20a, and mapped closest to L20a. The LOD score for the QTL was 21.8, and it explained 41% of the variation in the cross (Table 4-1). There were significant dominant and additive gene effects in this interval (Fig. 4-4a). In plants homozygous for the L20a allele from PI 594427C, the gall number was reduced by 86 galls compared to plants that were homozygous for the CNS allele at L20a (Table 4-2). The QTL in the population PI 594651L2 x CNS was mapped to the 0.8-cM interval between L20a and OPN11 (Fig. 4-4b), had a LOD score of 5.7, and explained 12% of the variation within that population (Table 4-1). The gene action underlying this QTL was additive, as the dominance effect was non-existent in the interval from L20a to OPN11 (Fig. 4-4b). Plants containing homozygous alleles from PI 594651L2 allele at OPN11 had approximately 43 fewer galls than the plants containing the CNS allele in homozygous condition at this locus (Table 4-2). No major QTLs were found in the population of PI 594403 x CNS, although the progeny segregated for the trait and all 188 individuals were genotyped with almost 200 SSR markers. Instead only minor QTLs were discovered, none of which explained more than 5% of the phenotypic variation (Table 4-1). The minor QTL on LG-E in the population of PI 594427C x CNS could only be detected when L20a was fixed for homozygous alleles. Under those conditions, the minor QTL mapped closest to Satt369, had a LOD score of 4.8, and accounted for 7% of the variation (Table 4-1). Based on 5000 permutations of the data, a QTL
with a LOD score of this size is considered highly significant (P<0.001). When plants had a combination of homozygous alleles from PI 594427C at L20a on LG-F and homozygous alleles from PI 594427C at Satt369 on LG-E, they averaged approximately 36 fewer galls than plants that were homozygous for the CNS alleles at Satt369 (Table 4-2). This minor QTL on LG-E was within 12.9 cM of another QTL on LG-E detected in the PI 594651L2 x CNS population, but the QTL in the latter population only had a LOD score of 1.5 (Table 4-1). Plants homozygous for the PI 594651L2 alleles at Sat_376 averaged 20 fewer galls than those homozygous for the CNS alleles at this marker locus (Table 4-2). In all of the above cases, the positive allele was contributed by the resistant parent and not by CNS.

More than 150 F2:3 families from each of three populations of PI 594403 x CNS, PI 594427C x CNS, and PI 594651L2 x CNS were evaluated to confirm the QTLs detected in the F2 populations. In the PI 594427C x CNS population, data were collected on 134 markers, while data were collected on 127 markers in the PI 594651L2 x CNS population. In the F2:3 population of PI 594403 x CNS, a total of 139 SSR markers were analyzed. The major QTLs on LG-F in the PI 594427C x CNS and PI 594651L2 x CNS populations were confirmed in similar positions in both populations. The QTL in the F2:3 population of PI 594427C x CNS was mapped within the interval from 83I9 to L20a, and accounted for 32% of the phenotypic variation in that population (Table 4-2). Similarly to the results in the F2 population of the same cross, there was evidence for both dominant and additive gene effects in this interval (data not shown). In the homozygous condition, the L20a allele from PI 594427C resulted in plants with gall ratings that averaged a score of 1.1 less than plants with the CNS allele at L20a (Table 4-2). The QTL detected in this F2:3 population is also similar in both position and magnitude of the R² value accounted for by the QTL flanked by 83I9 and L20a in the F2 population (Table 4-2). The F2:3
population of PI 594651L2 x CNS also contained a QTL on LG-F that was mapped to the interval between L20a and OPN11, had a LOD score of 6.4, and accounted for 17% of the variation in galling. This is the same position in which the QTL was mapped in the F2 population, and accounted for slightly more of the phenotypic variation than did the QTL in that population (Table 4-2). Similarly to the results in the F2 population, the gene action in this interval was estimated to be additive rather than dominant. Plants that contained the OPN11 allele from PI 594651L2 in the homozygous condition had a lower gall rating than those that contained the CNS allele at this locus in the homozygous condition (Table 4-2).

Only one minor QTL in the three F2 populations was confirmed in the F2:3 population. In the F2 population of PI 594651L2 x CNS, a QTL was detected on LG-E in the interval between Satt231 and Sat_376 with a LOD score of 1.5 and explaining 4% of the phenotypic variation. Although this accounts for only a small percentage of the variation, it was nonetheless confirmed in the F2:3 population of PI 594651L2 x CNS in the interval between Satt231 and Sat_376, with a LOD score of 2, and accounting for 6% of the variation. In this population, the presence of homozygous alleles from PI 594651L2 at Sat_376 resulted in plants with a lower gall rating than plants with the CNS allele at Sat_376 in the homozygous condition (Table 4-3). In contrast, the QTL on LG-E that was most closely associated with Satt369 was not confirmed in the F2:3 population of PI 594427C x CNS. In both F2:3 populations of PI 594427C x CNS and PI 594651L2 x CNS, minor QTLs were identified that had not been identified in the F2 populations of the same crosses. In each case, the minor QTLs accounted for less than 4% of the phenotypic variation, had LOD scores of less than 1.6, and were considered false positives.

Although more replications and plants were used in this confirmation experiment, no major QTLs were detected in the PI 594403 x CNS F2:3 population. Instead, minor QTLs, each
explaining less than 6% of the phenotypic variation, were detected. Since these were not the same minor QTLs that were detected in the F2 population, and since none of these exceeded the threshold of statistical significance based on permutation tests, it was assumed that these QTLs were false positives. Therefore, the real QTL(s) in PI 594403 still elude detection. Related to this are observations that PI 594403, although initially detected as a highly resistant germplasm source, had sustained abnormally high gall numbers in multiple greenhouse experiments that were conducted during the same time frame as was the phenotyping of these mapping populations (Yates, 2006). This seed source may be heterogeneous for Ma resistance, or, alternatively, may have been misidentified.

A resistance gene candidate on LG-F, L20a (Hayes et al., 2000), has been used to test for associations between its allelic variants and Ma resistance in over 60 different lines and cultivars. In almost every case, there is a close association between a particular allele pattern at this locus and Ma resistance (Yates, 2006). In this test, PI 200538, PI 230977, breeding lines developed with sources of Ma and Mj resistance from these two plant introductions, PI 594427C, and PI 594651L2, contain the L20a allele associated with Ma resistance. Additionally, almost every line or cultivar that was tested that had Ma resistance also contains the same L20a allele pattern. In contrast, PI 594403 and CNS contain the allele associated with Ma susceptibility (Fig. 4-5). This result suggests that PI 594403 does not contain the major Ma resistance allele on LG-F. Based on mapping results in this study, PI 594403 does not contain the Ma resistance allele at the minor QTL on LG-E either.

It appears that the resistance alleles on LG-F in PI 594651L2 and PI 594427C map to the same position, and may therefore represent either the same allele or different alleles that belong to the same locus. They also map close to the RFLP (restriction fragment length polymorphism)
marker B212, which was originally associated with the QTLs in PI 200538 and PI 230977 (Tamulonis et al., 1997a; Tamulonis et al., 1997b). According to a published fine-scale map, OPN11 and B212 map to the same position, but are approximately 0.4 cM from NBS5, which was later determined to be L20a (Gore et al., 2002; Hayes et al., 2000). The resistance allele(s) in PI 594427C and PI 594651L2 show differences in the mode of gene action. While the PI 594427C resistance allele expresses dominance, the PI 594651L2 resistance allele acts in an additive fashion. The resistance allele in PI 200538 also expresses dominance (Tamulonis et al., 1997b). This suggests that the alleles in PI 200538 and PI 594427C may be more similar to each other than to the allele from PI 594651L2. The resistance allele on LG-F from PI 594651L2 also did not reduce galling to as great of an extent. PI 594651L2 is a single-plant selection from within a susceptible plant introduction. It was selected for its unusual Ma response, as its roots did not sustain extensive galling but did contain high numbers of egg masses. This is unusual in that the typical response is for the numbers of egg masses and galls to be positively correlated. It is possible, therefore, that in phenotyping the progeny from these populations, egg masses were sometimes mistaken for galls, or that secondary Ma infection resulted in more galls on plants homozygous for the PI 594651L2 allele at this QTL. The overall effect would therefore lower the quality of the phenotypic data and the correlation coefficient in the marker-trait association. Alternatively, the resistance allele in PI 594651L2 may be unique from that in PI 594427C; one possibility is that its resistance is not expressed to the same extent as the allele in PI 594427C.

A minor QTL on LG-E was detected in the F2 population of PI 594651L2 x CNS and confirmed in the F2:3 population. PI 594651L2 was the donor of the resistance allele at this locus, which mapped closest to Sat_376. A minor QTL on LG-E was also discovered in the F2,
but not the $F_{2:3}$ population of PI 594427C x CNS. This QTL could only be detected in plants that were fixed for homozygous L20a alleles from PI 594427C, and mapped closest to Satt369. On the soybean genetic map, Sat_376 is separated from Satt369 by approximately 12.9 cM, therefore these two QTLs could represent the same locus. A minor QTL conditioning Ma resistance was previously mapped in PI 200538 with RFLP markers within a 31-cM region that includes Satt369, and is very close to Sat_376 on LG-E (Tamulonis et al., 1997b). Therefore, all three QTLs may represent the same allele or different alleles at the same locus.

The resistance alleles on LG-F in PI 594427C and PI 594651L2 not only map to similar positions to one another, but also coincide with the position of major resistance alleles in the Ma-resistant accessions in PI 200538 and PI 230977 (Tamulonis et al., 1997a; Tamulonis et al., 1997b). The 10-cM region of LG-F around the region of this QTL has been associated with resistance to multiple pests and pathogens, including both Ma and Mj (Tamulonis et al., 1997a; Tamulonis et al., 1997b), viruses (Fasoula et al., 2003; Yu et al., 1994), insects (Rector et al., 1999), fungi (Arahana et al., 2001), oomycete fungi (Burnham et al., 2003; Diers et al., 1992), and bacteria (Ashfield et al., 1998). As mentioned earlier, the Ma resistance response is associated with variation in the resistance gene candidate L20a, and variation at this locus almost perfectly coincides with Ma response in over 60 different cultivars, breeding lines, and PIs (unpublished results). Notably, PI 594403 contains the L20a allele variant that is associated with Ma susceptibility (Fig. 4-5). The genes conferring the Ma resistance in this PI were not identified in this study despite extensive marker testing, and this seed source may require either seed purification or further testing with different Ma isolates. Future directions include testing the association of L20a variation with the corresponding Ma reaction, and the development of L20a into a high-throughput marker such as a SNP (single nucleotide polymorphism) marker,
which will allow soybean breeders to easily determine which lines contain this important source of Ma resistance.

Acknowledgements

The authors gratefully acknowledge greenhouse assistance from Charles Lester, Matthew Ephgrave, Earl Baxter, Matt Wood, Tasaday Lynch, Renee Hilton, and Erika Swinson. This work was supported by funding from the United Soybean Board, the Georgia Agricultural Experiment Stations, and an NSF pre-doctoral fellowship to J.L. Yates.

References


Fig. 4-1. Rating scale used to score the F$_{23}$ families for extent of *M. arenaria* galling, where 1 represents plants with few galls and 5 represents plants with extensive galling.
Fig. 4-2. Frequency distribution of F$_2$ plants based on the number of $M$. arenaria galls in a) PI 594403 x CNS, b) PI 594427C x CNS, and c) PI 594651L2 x CNS.
Fig. 4-3. Frequency distributions of F_{2:3} lines based on *M. arenaria* gall ratings in a) PI 594403 x CNS, b) PI 594427C x CNS, and c) PI 594651L2 x CNS
Fig. 4-4. Interval mapping results for the *M. arenaria* gall number QTL detected on LG-F in the a) PI 594427C x CNS F₂ population and b) PI 594651L2 x CNS F₂ population, where —— indicates the LOD score, ———— indicates additive gene effects, and ———— indicates dominant gene effects. The ——— line indicates the LOD score which corresponds to a significance level of P<0.001 (a) and P<0.01 (b) determined from permutation tests.
Fig. 4-5. L20a allele patterns in (from L to R) CNS, PI 594403, PI 594427C, and PI 594651L2.
Table 4-1. Markers associated with *M. arenaria* galling in the F2 populations of PI 594403 x CNS, PI 594427C x CNS, and PI 594651L2 x CNS.

<table>
<thead>
<tr>
<th>Linkage Group</th>
<th>Marker</th>
<th>Distance (cM)</th>
<th>PI 594403 x CNS</th>
<th>PI 594427C x CNS</th>
<th>PI 594651L2 x CNS</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>LOD†</td>
<td>R²</td>
<td>LOD†</td>
<td>R²</td>
</tr>
<tr>
<td>A2</td>
<td>Sat_377</td>
<td>116.6</td>
<td>1.8</td>
<td>4</td>
<td></td>
</tr>
<tr>
<td>C1</td>
<td>Satt195</td>
<td>84.8</td>
<td>2</td>
<td>5</td>
<td></td>
</tr>
<tr>
<td>C2</td>
<td>Satt305</td>
<td>69.7</td>
<td>2.1</td>
<td>6</td>
<td></td>
</tr>
<tr>
<td>D1a</td>
<td>Satt071</td>
<td>20.5</td>
<td></td>
<td></td>
<td>2.7</td>
</tr>
<tr>
<td>D2</td>
<td>Satt135</td>
<td>26.0</td>
<td>2.2</td>
<td>5</td>
<td></td>
</tr>
<tr>
<td>E</td>
<td>Satt369</td>
<td>56.3</td>
<td>4.8</td>
<td>7</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Sat_376</td>
<td>69.2</td>
<td></td>
<td>1.5</td>
<td>4</td>
</tr>
<tr>
<td>F</td>
<td>OPN11</td>
<td>68.8</td>
<td></td>
<td>5.7</td>
<td>12</td>
</tr>
<tr>
<td></td>
<td>L20a</td>
<td>71.4</td>
<td></td>
<td>21.8</td>
<td>41</td>
</tr>
<tr>
<td>I</td>
<td>Sct_189</td>
<td>113.8</td>
<td></td>
<td>2.3</td>
<td>5</td>
</tr>
<tr>
<td>M</td>
<td>Satt567</td>
<td>33.5</td>
<td>2.1</td>
<td>5</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Satt 536</td>
<td>62.1</td>
<td></td>
<td>1.3</td>
<td>3</td>
</tr>
</tbody>
</table>

†LOD score for establishing significance at P<0.05 based on 5000 permutations: 3.6

‡LOD score for establishing significance at P<0.05 based on 5000 permutations: 3.5
Table 4-2. Description of LG-F and LG-E QTLs conditioning *M. arenaria* galling in F2 and F2:3 populations of PI 594427C x CNS and PI 594651L2 x CNS.

<table>
<thead>
<tr>
<th>Population</th>
<th>Linkage group</th>
<th>Marker</th>
<th>LOD</th>
<th>$R^2$</th>
<th>$2a^{\dagger}$</th>
<th>Positive allele</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>%</td>
<td>galls/gall rating‡</td>
<td></td>
</tr>
<tr>
<td>PI 594427CxCNS F2</td>
<td>F</td>
<td>L20a</td>
<td>21.8</td>
<td>41</td>
<td>86</td>
<td>PI 594427C</td>
</tr>
<tr>
<td>PI 594427CxCNS F2</td>
<td>E</td>
<td>Satt369</td>
<td>4.8</td>
<td>7</td>
<td>36</td>
<td>PI 594427C</td>
</tr>
<tr>
<td>PI 594427CxCNS F3</td>
<td>F</td>
<td>L20a</td>
<td>13.8</td>
<td>32</td>
<td>1.1</td>
<td>PI 594427C</td>
</tr>
<tr>
<td>PI 594651L2xCNS F2</td>
<td>F</td>
<td>OPN11</td>
<td>5.7</td>
<td>12</td>
<td>43</td>
<td>PI 594651L2</td>
</tr>
<tr>
<td>PI 594651L2xCNS F3</td>
<td>F</td>
<td>OPN11</td>
<td>6.4</td>
<td>17</td>
<td>0.6</td>
<td>PI 594651L2</td>
</tr>
<tr>
<td>PI 594651L2xCNS F2</td>
<td>E</td>
<td>Sat_376</td>
<td>1.5</td>
<td>4</td>
<td>21</td>
<td>PI 594651L2</td>
</tr>
<tr>
<td>PI 594651L2xCNS F3</td>
<td>E</td>
<td>Sat_376</td>
<td>2</td>
<td>6</td>
<td>0.2</td>
<td>PI 594651L2</td>
</tr>
</tbody>
</table>

$^{\dagger}2a= \text{Difference in number of galls or in galling score between plants that are homozygous for the PI alleles versus plants that are homozygous for the CNS alleles at the indicated marker locus.}$

$^{\ddagger}\text{gall rating}= \text{rating scale from 1 to 5, where 1 indicates plant with very few galls from } M.\ arenaria \text{ parasitism, and 5 indicates plants with many galls}$
CHAPTER 5

MOLECULAR PEDIGREE ANALYSIS OF A MAJOR QTL CONDITIONING RESISTANCE TO PEANUT ROOT-KNOT NEMATODES IN SOYBEAN

*Yates, J.L, R.S. Hussey, S.L. Finnerty, and H.R. Børma. To be submitted to Crop Sci.*
Abstract

Root-knot nematodes (RKN) are of substantial economic importance as a result of their ability to parasitize and reduce yields of all major crops grown worldwide. In soybean fields in Georgia, the peanut RKN (*Meloidogyne arenaria* (Neal) Chitwood or Ma) is increasingly a problem both because of its ability to parasitize soybean [*Glycine max* (L.) Merr.] and peanut (*Arachis hypogaea* L.), and because populations of other RKN species have traditionally been targeted. A major quantitative trait locus (QTL) conditioning resistance to Ma was previously mapped to the same relative position on linkage group (LG) F of the soybean genetic map in at least three Ma-resistant germplasm sources of diverse origins (Tamulonis et al., 1997b; Yates, 2006), and thus appears to be an important source of Ma resistance. The QTL in all three sources was closely associated with the markers L20a and OPN11. In this study, 61 lines, plant introductions (PIs) and cultivars were evaluated for Ma response as well as allele composition at L20a and OPN11. The allelic variation at L20a was highly associated with the published and the empirically determined Ma reaction in 61 soybean PIs, lines, and cultivars. The allelic variation in OPN11 did not perfectly correlate with the Ma response, and instead showed possible evidence in two lineages of linkage in repulsion phase with the resistance gene. Variation at OPN11, however, was informative in determining that ‘Palmetto’ was the source of Ma resistance in many cultivars commonly used as parents in southern soybean breeding programs, such as ‘Bragg’, ‘Forrest’, and ‘Braxton’. Since L20a corresponds to a resistance gene candidate, it is possible that this gene is either tightly linked to the Ma resistance gene or directly provides the Ma resistance itself. L20a allelic variation is thus highly predictive of Ma resistance and could be effectively used as part of a marker-assisted selection strategy to incorporate Ma resistance into soybean breeding lines.
**Introduction**

Root-knot nematodes cause more damage to the global food supply than other category of plant-parasitic nematode (Bernard, 1989). Although the genus contains over 50 described species, only four are of substantial economic importance, causing 95% of the incidences of RKN parasitism on crops worldwide (Hussey and Janssen, 2002). These four species, *Meloidogyne incognita* (Kofoid and White) Chitwood (Mi, southern RKN), *M. arenaria* (Neal) Chitwood (Ma, peanut RKN), *M. javanica* (Treub) Chitwood (Mj, Javanese RKN), and *M. hapla* Chitwood (Mh, northern RKN), are capable of parasitizing over 2000 plant species, with species representing almost every known plant family (Roberts, 1995; Trudgill and Blok, 2001). Mi and Ma present the most problems to soybean growers in the southeastern USA (Boerma and Hussey, 1992; Rodríguez-Kábana et al., 1991). While Ma is found within the same latitudinal distribution as Mi, it is not as common as either Mi or Mj. However, its morphology, biochemistry, and cytological profile are more variable and atypical than either Mi or Mj (Cliff and Hirschmann, 1985; Eisenback and Triantophyllou, 1991).

Since Mi resistance in soybean is not effective against Ma, the intensive cultivation of Mi-resistant cultivars of tobacco (*Nicotiana tabacum* L.), cotton (*Gossypium* L.), and soybean has resulted in substantial increases in Ma population densities in the southeastern USA (Fortnum et al., 1984; Garcia and Rich, 1985; Rodríguez-Kábana and Williams, 1981; Schmitt and Barker, 1988). In addition, peanut is the preferred host of race 1 of Ma, and the hectareage of this crop has traditionally been substantial in the southern USA. Since Ma is not as sensitive to nematicides as Mi (Barker et al., 1981), and isolates can be more aggressive on soybean than those of Mi (Starr, 1989), Ma may be a more difficult pest to manage than Mi. As a result of the extensive variation within this species, the corresponding response to Ma in soybean has also
been quite variable with different Ma populations (Kinloch et al., 1987; Noe, 1992). Although sources of Ma resistance in soybean cultivars and plant introductions have been identified (Harris et al., 2003; Luzzi et al., 1987), characterized (Luzzi et al., 1995; Tamulonis et al., 1997b), and introgressed into breeding lines (Luzzi et al., 1996), the existing level of Ma resistance is not nearly as effective as that of Mi in soybean.

Further complicating the situation is the confusion between Ma tolerance and Ma resistance. For years, Ma was not considered a problem in soybean because many cultivars that were chosen for their Mi resistance were Ma-tolerant (Carpenter and Lewis, 1991; Ibrahim and Lewis, 1993; Kinloch et al., 1987), meaning that Ma was capable of reproducing on the cultivar but did not substantially reduce its yield. This benefits the grower in that yield is not reduced, but becomes problematic in the following seasons, particularly if Ma-susceptible and/or Ma-intolerant crops are planted. In reality, few soybean cultivars possess high levels of Ma resistance (Garcia and Rich, 1985), and cultivars once considered Ma resistant were later re-characterized as Ma tolerant (Hiatt III et al., 1988).

This study was initiated to assess both the level of Ma resistance in southern U.S. soybean cultivars and their ancestors, and to determine the ancestral source of the Ma resistance through a DNA marker pedigree analysis. Such a study is feasible since southern U.S. soybean cultivars derive approximately 73 to 80% of their parentage from fewer than seven ancestral lines (Delannay et al., 1983; Sneller, 1994). Ha et al. (2004) used this approach to find an almost perfect association between Mi resistance and particular allele sizes at two SSR (simple sequence repeat) marker loci on linkage group (LG) O in 48 soybean genotypes with common ancestors. Similarly, Narvel et al. (2001) used SSRs to determine that at least 13 of 15 insect-resistant
soybean lines or cultivars sharing a common insect-resistant donor parent contained the same SSR marker bands at a QTL on LG-M of the genetic map of soybean.

The source of the Ma resistance in these elite southern cultivars and their ancestors can be determined in this study by using markers that are associated with a major locus associated with Ma resistance on LG-F (Tamulonis et al., 1997b). This 10-cM region of LG-F is a source of genes conferring resistance to multiple pests and pathogens, including both Ma and Mj (Tamulonis et al., 1997a; Tamulonis et al., 1997b), viruses (Fasoula et al., 2003; Yu et al., 1994), insects (Rector et al., 1999), fungi (Arahana et al., 2001; Burnham et al., 2003; Diers et al., 1992), and bacteria (Ashfield et al., 1998). Its importance as a major locus for Ma resistance has been established in three different soybean plant introductions (PIs) of diverse origins (Tamulonis et al., 1997b; Yates, 2006), making it a logical choice to examine in the pedigrees of these lines. L20a and OPN11 are markers developed from a resistance gene candidate and from a RAPD (random amplified polymorphic DNA) fragment, respectively, and were previously reported to be associated with resistance genes (Hayes et al., 2000; Zheng et al., 2003). Both of these markers were closely associated with the QTL conditioning Ma response in three different PIs (Yates, 2006). This study evaluates the association of the allelic variation at L20a and OPN11 with the Ma resistance reaction in southern soybean cultivars and their ancestors.

Materials and Methods

Seeds of 61 soybean genotypes representing released cultivars, breeding lines, and ancestral plant introductions were obtained from the USDA Soybean Germplasm Collection (Univ. of Illinois, Urbana, IL) (Table 5-1). The soybean genotypes were chosen both for their known or suspected reaction to Ma and for their known pedigree relationships. Each genotype was evaluated for its Ma resistance in a randomized complete block design with six replications.
Three seeds of each genotype were individually planted in 20.6-cm Ray Leach single-cell conetainers (Stuewe & Sons, Inc., Corvallis, OR), filled with a soil mix, and placed in every other column of Ray Leach 98-cone trays. The soil mix consisted of Pacolet sandy loam, which belongs to the fine, kaolinitic, thermic family of Typic Kanhapludults. The sandy loam was supplemented with sand to a final weight composition of 73% sand, 16% silt, and 11% clay. An extra tray of checks of known Ma response were included both to establish a scale for evaluating the genotypes and for determining the optimal time to rate the experiment. The checks included G93-9106 (highly Ma resistant), ‘Haskell-RR’ (resistant), ‘Boggs-RR’ (moderately resistant), ‘Benning-RR’ (susceptible), ‘Prichard-RR’ (susceptible), and ‘CNS’ (highly susceptible).

A week after planting, two of the three seedlings in each cone were removed to leave one plant per cone. The genotypes were inoculated with 3000 Ma eggs, which were dispensed using a digital dispensing pump (Asteck Systems Corp., Farmingdale, NY). The eggs were obtained from the root systems of the susceptible eggplant (Solanum melongena L.) cultivar Black Beauty, and were collected with 0.5% NaOCl as described previously (Hussey and Janssen, 2002). The plants were kept in a greenhouse maintained at approximately 28°C and were watered manually three times daily. Once a week the plants were watered with a solution that was supplemented with 6 mg N, 3 mg P, and 5 mg K. During the last 2 wk of growth, the plants were moved to another greenhouse with an automated overhead irrigation system, where they were watered three times daily.

When a significant level of galling had developed on the roots of CNS, the experiment was terminated. The tops of the plants were removed, and the root systems were removed from the cones, rinsed free of soil debris, and examined for galling. The checks were used to develop
a rating scale for galling, where 1 represents root systems with little to no evidence of galling and 5 represents extensive galling as would be typical of that of CNS or Prichard-RR.

Approximately 2 wk after planting, a trifoliolate leaf was removed from each plant in each replication to be used for genomic DNA extraction. The six leaves per genotype were lyophilized in a freeze-dryer for approximately 48 hours, and were then placed in individual wells of two 48-well, deep-well plates along with a zinc-plated 4.5-mm BB (Daisy Outdoor Products, Rogers, AR). Each plate was placed in a custom-made leaf grinder and shaken (exact RPM unknown) for approximately 1 min to pulverize the tissue. DNA extraction then proceeded as described in Keim et al. (1988). The DNA pellet was resuspended in TE buffer (10 mM Tris-HCl, 1 mM EDTA pH 8) and adjusted to a concentration of approximately 50 ng µL⁻¹ for use in PCR (polymerase chain reaction).

The primers used in this experiment were designed from sequences available in this interval of LG-F between the SSR markers SoyHsp176 and Satt510 (Gore et al., 2002). L20a and OPN11 are each described in previous reports (Hayes et al., 2000; Zheng et al., 2003), although the primer sequences used to amplify the L20a segment were not described by Hayes et al. (2000). The L20a primer sequences designed and used in this study are F-5’-TGT GAG GCA TCA GAA GGG G-3’ and R-5’-GCC ACC TAT GCC ACA AAT CC-3’. The PCR conditions for OPN11 are described elsewhere (Zheng et al., 2003), while for L20a the PCR mix was a 25-µL volume consisting of 1X buffer (10 mM Tris-HCl pH 9.0, 50 mM KCl, 0.1% Triton®X-100), 2.5 mM MgCl₂, 0.2 mM dNTPs, 1 unit Taq DNA polymerase (M1668, Promega Corp., Madison, WI), and approximately 50 ng µL⁻¹ DNA template. The PCR cycling conditions consisted of 30 cycles of a denaturation step at 94°C for 30 s, 54°C for 30 s, and 72°C for 1 min. Approximately 3 µL of the PCR products from OPN11 and L20a were combined with 9 µL of a loading dye.
solution (950 µL deionized formamide, 40 µL 1% of each of bromophenyl blue and xylene cyanol, and 10 µL 1 M NaOH) and denatured for 2 min at 95°C. Since the PCR products among the parents did not contain length polymorphisms, sequence differences were detected instead using the SSCP (single-strand conformation polymorphism) technique. The samples were loaded on a 0.5X MDE™ (Cambrex Bio Science, Rockland, Inc.) gel (71.5 mL H₂O, 27.5 mL 2X MDE™, 11 mL 6X TBE, 800 µL 10% APS, 160 µL TEMED) on a vertical MegaGel polyacrylamide gel apparatus (C.B.S. Scientific Company, Solana Beach, CA), and electrophoresed at 100 V (1 W gel⁻¹) for approximately 40 hours. At the end of the run, the PCR products were detected by silver staining, first by agitating in fixer (210 mL 95% ethanol and 10 mL glacial acetic acid brought to a 1-L volume) for 10 min, followed by 10 min in silver nitrate solution (2% silver nitrate in fixer), two quick washes with deionized H₂O, and treatment with developer (75 mL 10 N NaOH, 1 mL 37% formaldehyde brought to a 1-L volume) until the PCR fragments were at the desired intensity. The gel was then rinsed once with fixer to stop the reaction and was photographed.

**Results and Discussion**

The Ma reaction of most of the lines or cultivars tested in this study was reported either in the cultivar release notice or in the Southern Regional Uniform Soybean Tests (Table 5-1). However, prior to the mid-1970s, the particular root-knot nematode species used to assess resistance in the potential cultivars was not distinguished as Mi or Ma. Therefore, many cultivars released prior to this time were not specifically evaluated for Ma resistance. The Ma reactions of some of the original ancestors, such as PI 54610, are also not reported in the literature (Table 5-1).
In addition to the Ma responses that were reported in the cultivar release notices and the Southern Regional Uniform Tests, many of the same soybean cultivars have been characterized in different greenhouse, field, or microplot studies for their levels of Ma resistance. ‘Bedford’, ‘Bossier’, ‘Braxton’, ‘Bragg’, ‘Coker 6738’, ‘Forrest’, ‘Gordon’, ‘Govan’, ‘Lee’, ‘Perrin’, and ‘Wright’ show moderate to high levels of Ma resistance (Carpenter and Lewis, 1991; Dropkin, 1959; Hiatt III et al., 1988; Ibrahim and Lewis, 1993; Kinloch, 1980; Koenning and Barker, 1992; Rodríguez-Kábana and Thurlow, 1980). ‘Centennial’, ‘Cobb’, ‘Dowling’, ‘Gasoy17’, ‘Hutton’, and ‘Young’ are reportedly Ma-susceptible (Carpenter and Lewis, 1991; Hiatt III et al., 1988; Ibrahim and Lewis, 1993; Koenning and Barker, 1992; Rodríguez-Kábana and Thurlow, 1980). ‘Jackson’ was reported to be moderately galled in one study (Dropkin, 1959), yet it is often used as a resistant check in our greenhouse studies, showing additional evidence of the variable response of the same cultivar to different Ma isolates.

The Ma reaction determined in this study closely matched the Ma response that was reported in the literature. One exception was Forrest, which is reportedly Ma resistant (Hussey et al., 1991), but had a high level of galling in this study (Table 5-1). Another exception is Hutcheson, which was reported in the literature to be Ma susceptible, but showed a moderately low level of galling in this study. Based on our data there is a discrepancy for some genotypes that had mean gall numbers in the range of 50 to 55 galls and the Ma reaction reported in the literature (Table 5-1). As stated earlier, Ma is a variable nematode species, and it is common to observe variable galling results in the same line or cultivar in different experiments. This particular study was carried out for a slightly longer period of time than typical greenhouse studies with Ma. Since Ma takes from 25 to 30 d to complete its life cycle, the extended length of the experiment in this study could have supported at least two generations of Ma. It is
therefore possible that the resistance in certain lines or cultivars was overcome, resulting in root systems that have more galling than is normally observed for this line.

Genomic DNA extracted from the trifoliolate leaves of at least six plants per line or cultivar was analyzed with the PCR-based markers L20a and OPN11. These markers were chosen based on their close association with quantitative trait loci (QTLs) conditioning Ma resistance detected in at least two Ma-resistant plant introductions (Tamulonis et al., 1997b; Yates, 2006).

The allelic variation in the marker L20a, the primers of which were developed from the nucleotide sequence of a resistance gene candidate, was highly associated with the published Ma response and with the Ma reaction assessed in the greenhouse in this study (Table 5-1). For example, L20a allelic variation in seven cultivars, Centennial, Cobb, Coker338, Braxton, Gordon, Govan, and Young is shown in Figure 5-1. The Ma-susceptible cultivars Young, Centennial, Cobb, and Coker338 contain an L20a allele, which appears as a separated set of bands on the SSCP gel (allele “1”, Fig. 5-1), that is highly associated with Ma susceptibility (Fig. 5-2). With L20a, the Ma-resistant cultivars Braxton, Gordon, and Govan had an allele, which on the SSCP gel appeared as a set of bands that were closer together than the other set of bands (allele “2”, Fig. 5-1), that was associated in almost every instance with Ma resistance in both the major S-100 x CNS and (Tokyo x PI 54610) x Palmetto lineages (Fig. 5-2). There were few exceptions to this relationship. Pharaoh and Pickett71 had allele “2”, which is associated with Ma resistance; however, both cultivars are reported in the literature to be Ma susceptible, and both averaged either 60 or more Ma galls in this study (Table 5-1). The Ma response of PI54610 had not been previously reported in the literature; in this study, PI54610 showed a similar level of galling as Pharoah and Pickett71, yet it also had allele “2” at the L20a locus. Lastly, Benning
was reported in the literature to be moderately resistant, and it showed a fairly low level of
galling in this study, but it had the allele pattern that is associated with Ma susceptibility (allele
“1”, Table 5-1). These inconsistencies could be the result of recombination events between L20a
and the resistance allele, or could be due to a misclassification of these cultivars in the literature
as resistant or susceptible. This is possible since Ma is a variable species, and different Ma
isolates can differ widely in their aggressiveness. Cultivars can also respond in an equally
variable manner in different greenhouse experiments depending on the length of the experiment
and the aggressiveness of the Ma isolate.

The association of Ma resistance with L20a allelic variation reveals that Palmetto and
S-100 are donors of the Ma resistance that is present in the lines and cultivars that arose from
these ancestors (Fig. 5-2). In addition, the L20a allele “2” is also present in all known
ergermplasm sources of Ma resistance, including PI 230977, PI 200538, PI 594427C, and PI
594651L2 (Yates, 2006). The tight association between Ma response and allelic variation makes
L20a a useful marker for soybean breeders to select for Ma resistance in their cultivar
development programs.

OPN11 was also tested to determine whether variation at another closely linked marker
would also co-segregate with Ma response. OPN11 corresponds to a non-genic region that is
tightly linked to the \( Rsv1 \) gene that confers resistance to the soybean mosaic virus (Zheng et al.,
2003). The PCR products from OPN11 showed three main sets of bands on the SSCP gel, one
set consisting of two lower bands (allele “2”), one set consisting of two upper bands (allele “1”),
and one set consisting of a single top band (allele “3”, Fig. 5-3). CNS had an additional fourth
set of bands (data not shown), but this allele was not found in any of its descendants. In Figure
5-3, the OPN11 allelic variation among the cultivars matches the expected Ma response with one
exception, which is Gordon. Gordon is Ma-resistant, but has the same allele at OPN11 as the Ma-susceptible cultivars Cobb, Coker 338, and Young. This instance in Gordon in which the Ma response does not match the marker pattern was found to occur throughout the S-100 x CNS lineage, beginning with S-100 itself (Ma resistant and allele “1”, Fig. 5-4, Table 5-1). The opposite Ma reaction/OPN11 allele configuration, that is, Ma susceptibility paired with OPN11 allele “2”, was discovered in Sharkey and Lamar (Fig. 5-4, Table 5-1). These results indicate a different linkage phase for the S-100 lineage (Ma resistant and allele “1”) and the Palmetto lineage (Ma resistant and allele “2”). In Figure 5-3, Govan and Braxton are Ma-resistant and have allele “2” that is associated with Ma resistance throughout the Palmetto x Volstate lineage (Fig. 5-4). Allele “3” amplified in Centennial was not as common as the others, but was the same as that in PI 54610 (Fig. 5-4). Since the PI 54610 seed source used in this study is likely not representative of the original seed source used to create Ogden and Volstate, it is difficult to draw any conclusions about the association of this allele with Ma reaction, although in many cases, such as with Centennial and Volstate, it appears to be associated with Ma susceptibility (Table 5-1). Interestingly, it is heterogeneous with the Palmetto resistance allele in the Ma-resistant cultivars Jackson and Bryan (Fig. 5-4). As mentioned earlier, CNS had an OPN11 allele that was not observed in any of its descendants (allele “4”, Fig. 5-4, Table 5-1). Overall, the allelic variation suggests that OPN11 is tightly linked to the Ma resistance gene(s) but there is some degree of recombination between this locus and the Ma resistance gene. Variation at this locus is not perfectly predictive of Ma response, therefore OPN11 would not be the ideal marker to use to select for Ma-resistant lines. An exception would be if the breeder knows the linkage phase of OPN11 with the resistance allele, for example, whether the line obtained the OPN allele originally from Haskell (Ma resistant, allele “2”) or from Gordon (Ma resistant, allele “1”).
OPN11 allelic variation is also informative with regards to the source of the resistance alleles throughout the lineages. For example, it is clear from Figure 5-4 that Palmetto is the source of Ma resistance in Bragg, Forrest, and Braxton, which are the parents of many modern elite soybean cultivars that are currently grown in the southeastern USA.

It is intriguing that L20a is a resistance gene candidate of unknown pathogen specificity, particularly since it has such a close association with Ma reaction. In attempting to determine its specificity, Hayes et al. (2000) evaluated but, unlike the present study, found no association between allelic variation at L20a and response among genotypes that were resistant or susceptible to various pests and pathogens. In its previous designation as NBS5, L20a was mapped only 0.4 cM apart from OPN11 (Gore et al., 2002). Although there is no simple way to prove that L20a is the resistance gene responsible for the Ma resistance phenotype without cloning and complementation studies, the fact that two such tightly linked markers do not both co-segregate with Ma reaction seems to indicate that L20a is either more tightly linked to the Ma resistance gene than OPN11, or that L20a could be the Ma resistance gene itself. The use of the L20a marker or a single nucleotide polymorphism (SNP) marker developed from L20a, in cultivar development programs would ensure that the level of Ma resistance already present in the pedigrees of southern soybean cultivars is maintained.

Acknowledgements

The authors gratefully acknowledge greenhouse assistance from Charles Lester, Matt Wood, and Erika Swinson. This work was supported by funding from the United Soybean Board, the Georgia Agricultural Experiment Stations, and an NSF pre-doctoral fellowship to J.L. Yates.

References


Fig. 5-1 Sequence variation in L20a PCR products on a SSCP gel in (from L to R): Young (susceptible to *M. arenaria* or S), Braxton (resistant to *M. arenaria* or R), Centennial (S), Cobb (S), Coker338 (S), Gordon (R), and Govan (R).
Fig. 5-2. The relationship of *M. arenaria* (Ma) reaction to L20a allelic variation. The shaded boxes indicate Ma susceptibility, and the open boxes indicate Ma resistance as determined in greenhouse tests in this study. The different L20a alleles are designated by “1” and “2”. The “1/2” in Dyer indicates that it is heterogeneous at this locus. Solid lines indicate a direct relationship between the cultivars or lines, while dashed lines represent an indirect relationship.
Fig. 5-3. Sequence variation in OPN11 PCR products on a SSCP gel in (from L to R): Young (susceptible to *M. arenaria* or S), Braxton (resistant to *M. arenaria* or R), Centennial (S), Cobb (S), Coker338 (S), Gordon (R), and Govan (R).
Fig. 5-4. The relationship of *M. arenaria* (Ma) reaction to OPN11 allelic variation. The shaded boxes indicate Ma susceptibility, and the open boxes indicate Ma resistance. The different OPN11 alleles are designated by numbers 1 to 4. Cultivars that are heterogeneous for two alleles have two numbers separated by a forward slash. Solid lines indicate a direct relationship between the cultivars or lines, while dashed lines represent an indirect relationship.
Table 5-1. *M. arenaria* (Ma) gall number, published Ma reaction, allelic composition at the L20a and OPN11 loci, and pedigree information for 61 soybean genotypes.

<table>
<thead>
<tr>
<th>Genotype</th>
<th>Mean galls plant&lt;sup&gt;1&lt;/sup&gt;</th>
<th>Published Ma reaction&lt;sup&gt;†&lt;/sup&gt;</th>
<th>Allele at L20a&lt;sup&gt;‡&lt;/sup&gt;</th>
<th>Allele at OPN11&lt;sup&gt;‡&lt;/sup&gt;</th>
<th>Pedigree</th>
<th>Year of release</th>
<th>Source</th>
</tr>
</thead>
<tbody>
<tr>
<td>Perrin</td>
<td>12</td>
<td>R</td>
<td>2</td>
<td>2</td>
<td>Coker 488 × Braxton</td>
<td>1988</td>
<td>Shipe et al. (1990)</td>
</tr>
<tr>
<td>Coker 6738</td>
<td>14</td>
<td>MR</td>
<td>2</td>
<td>2</td>
<td>Braxton × Coker 368</td>
<td>1987</td>
<td>Hussey et al. (1991)</td>
</tr>
<tr>
<td>Colquitt</td>
<td>18</td>
<td>MR</td>
<td>2</td>
<td>2</td>
<td>Wright × Braxton</td>
<td>1990</td>
<td>Boerma et al. (1990)</td>
</tr>
<tr>
<td>Gordon</td>
<td>19</td>
<td>MR</td>
<td>2</td>
<td>1</td>
<td>Forrest × Pickett 71</td>
<td>1984</td>
<td>Boerma et al. (1985)</td>
</tr>
<tr>
<td>Bragg</td>
<td>22</td>
<td>R</td>
<td>2</td>
<td>2</td>
<td>Jackson × D49-2491 (S-100×CNS)</td>
<td>1963</td>
<td>Hartwig and Jamison (1975)</td>
</tr>
<tr>
<td>Govan</td>
<td>22</td>
<td>R</td>
<td>2</td>
<td>2</td>
<td>Bragg × Semmes</td>
<td>1977</td>
<td>Hartwig et al. (1978)</td>
</tr>
<tr>
<td>Wright</td>
<td>23</td>
<td>MR</td>
<td>2</td>
<td>2</td>
<td>Bragg × Lee</td>
<td>1979</td>
<td>Boerma et al. (1980)</td>
</tr>
<tr>
<td>Gregg</td>
<td>23</td>
<td>R</td>
<td>2</td>
<td>1</td>
<td>Bragg × Pickett 71</td>
<td>1983</td>
<td>Hartwig and Edwards (1983)</td>
</tr>
<tr>
<td>Twiggs</td>
<td>24</td>
<td>MR</td>
<td>2</td>
<td>2</td>
<td>Centennial × Forrest</td>
<td>1988</td>
<td>Boerma et al. (1988)</td>
</tr>
<tr>
<td>D49-2491</td>
<td>25</td>
<td>?</td>
<td>2</td>
<td>1</td>
<td>S-100×CNS</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>Manokin</td>
<td>26</td>
<td>R</td>
<td>2</td>
<td>1 and 2</td>
<td>L70L-3408 × D74-7824 (Forrest × D70-3001)</td>
<td>1991</td>
<td>Kenworthy et al. (1996)</td>
</tr>
<tr>
<td>Lee</td>
<td>28</td>
<td>R</td>
<td>2</td>
<td>1</td>
<td>S-100 × CNS</td>
<td>1954</td>
<td>Dropkin (1959)</td>
</tr>
<tr>
<td>Variety</td>
<td>Year</td>
<td>Source</td>
<td>Parents</td>
<td>Notes</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>-----------</td>
<td>------</td>
<td>--------</td>
<td>---------</td>
<td>-------</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Palmetto</td>
<td>1936</td>
<td>Luzzi et al. (1987)</td>
<td>Selected from PI 71587</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Braxton</td>
<td>1979</td>
<td>Hussey et al. (1991)</td>
<td>F59-1505 × (Bragg (3)†† × D60-7965)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Semmes</td>
<td>1966</td>
<td></td>
<td>D51-5427 (Ralsoy x Ogden) × D49-2491 (S-100 x CNS)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lee 74</td>
<td>1974</td>
<td></td>
<td>Lee 68 × R66-1517 (Lee (5) × FC 33.243)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bossier</td>
<td>1958</td>
<td>Hussey et al. (1991)</td>
<td>A late mutant in Lee</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Benning</td>
<td>1996</td>
<td>Boerma et al. (1997)</td>
<td>Hutcheson x Coker 6738</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Haskell</td>
<td>1993</td>
<td>Boerma et al. (1994b)</td>
<td>Johnston × Braxton</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bedford</td>
<td>1978</td>
<td>Hussey et al. (1991)</td>
<td>Forrest (2) × (D68-18 x PI88788)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Walters</td>
<td>1990</td>
<td>Caviness et al. (1991)</td>
<td>Forrest × Narow</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ogden</td>
<td>1953</td>
<td>Dropkin (1959)</td>
<td>Tokyo x PI 54610</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hutton</td>
<td>1972</td>
<td>Hartwig and Jamison (1975)</td>
<td>F55-822 (Jackson × D49-2491) × (Roanoke × CNS-4)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Jackson</td>
<td>1953</td>
<td>Hussey et al. (1991)</td>
<td>Volstate (2) × Palmetto</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>S-100</td>
<td>1945</td>
<td>Dropkin (1959)</td>
<td>Selected from 'Illini' in 1938 by Lee Mumford, farmer, Rutledge, Missouri. Reselected in 1942 by the Missouri AES</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Variety</td>
<td>Year</td>
<td>Location</td>
<td>Parentage/Description</td>
<td>Year</td>
<td>References</td>
<td></td>
<td></td>
</tr>
<tr>
<td>------------</td>
<td>------</td>
<td>----------</td>
<td>--------------------------------------------------------------------------------------</td>
<td>------</td>
<td>--------------------------------------</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pharaoh</td>
<td>60</td>
<td>S</td>
<td>2 2 Forrest × V71-480</td>
<td>1989</td>
<td>Schmidt et al. (1993)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lamar</td>
<td>61</td>
<td>S</td>
<td>1 2 Tracy M × selection (Centennial × D75-10169)</td>
<td>1989</td>
<td>Hussey et al. (1991)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ransom</td>
<td>61</td>
<td>S</td>
<td>1 1 (N55-5931 × N55-3818) × D56-1185</td>
<td>1970</td>
<td>Brim and Elledge (1973)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>PI 54610</td>
<td>61</td>
<td>?</td>
<td>2 3 'Pickett' × phytophthora rot-resistant 'Lee' type</td>
<td>-</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pickett71</td>
<td>62</td>
<td>S</td>
<td>2 1 'Pickett' × phytophthora rot-resistant 'Lee' type</td>
<td>1971</td>
<td>Hartwig and Jamison (1975)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>CNS</td>
<td>68</td>
<td>S</td>
<td>1 4 Selected from 'Clemson'</td>
<td>1943</td>
<td>Hussey et al. (1991)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Coker 338</td>
<td>68</td>
<td>S</td>
<td>1 1 Hampton 266 × Bragg</td>
<td>1973</td>
<td>Hartwig and Jamison (1975)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tokyo</td>
<td>72</td>
<td>?</td>
<td>1 1 Variety name assigned directly to or variety</td>
<td>1907</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PI 200492</td>
<td>73</td>
<td>?</td>
<td>1 1 Selected from PI 8424</td>
<td>-</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Dyer</td>
<td>74</td>
<td>?</td>
<td>1 and 2 1 Hill × (Lee (2) × Peking)</td>
<td>1967</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Dowling</td>
<td>74</td>
<td>S</td>
<td>1 1 Semmes × PI 200492</td>
<td>1978</td>
<td>Craigmiles et al. (1978)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Volstate</td>
<td>76</td>
<td>?</td>
<td>1 3 Tokyo × PI 54610</td>
<td>1942</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Forrest</td>
<td>77</td>
<td>MR</td>
<td>2 2 Dyer × Bragg</td>
<td>1972</td>
<td>Luzzi et al. (1987)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Johnston</td>
<td>78</td>
<td>S</td>
<td>1 1 N70-2173 (Hampton × Ransom) × Hutton</td>
<td>1983</td>
<td>Hussey et al. (1991)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hampton</td>
<td>79</td>
<td>?</td>
<td>1 1 Majos × Lee</td>
<td>1962</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Coker 488</td>
<td>79</td>
<td>S</td>
<td>1 1 Hampton 266 × Bragg</td>
<td>1977</td>
<td>Hussey et al. (1991)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cook</td>
<td>79</td>
<td>S</td>
<td>1 1 Braxton × Young</td>
<td>1991</td>
<td>Hartwig and Edwards (1989)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Dillon</td>
<td>81</td>
<td>S</td>
<td>1 1 Centennial × Young</td>
<td>1995</td>
<td>Hartwig and Kenty</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Variety</td>
<td>Year</td>
<td>Code</td>
<td>Co-variety 1</td>
<td>Co-variety 2</td>
<td>Notes</td>
<td></td>
<td></td>
</tr>
<tr>
<td>------------</td>
<td>------</td>
<td>------</td>
<td>--------------</td>
<td>--------------</td>
<td>-----------------------------------------------------------------------</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ralsoy</td>
<td>1992</td>
<td>84</td>
<td>?</td>
<td>1</td>
<td>Selected from 'Arksoy'</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Gasoy17</td>
<td>1977</td>
<td>85</td>
<td>S</td>
<td>1</td>
<td>Bragg × Hood</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Doles</td>
<td>1993</td>
<td>87</td>
<td>S</td>
<td>1</td>
<td>D74-7741 (Forrest × D70-3001) × Young</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hartwig</td>
<td>1991</td>
<td>88</td>
<td>S</td>
<td>3</td>
<td>Forrest (3) × PI 437654</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hampton</td>
<td>1991</td>
<td>89</td>
<td>?</td>
<td>1</td>
<td>Single-plant selection within Hampton: Majos × Lee</td>
<td></td>
<td></td>
</tr>
<tr>
<td>266A</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Maxey</td>
<td>1992</td>
<td>89</td>
<td>S</td>
<td>1</td>
<td>D76-9665 (Forrest × Centennial) × Johnston</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cobb</td>
<td>1973</td>
<td>90</td>
<td>S</td>
<td>1</td>
<td>F57-735×D58-3358 (Jackson (4) × D49-2491)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>FC 33.243-R-4</td>
<td>91</td>
<td>S</td>
<td>?</td>
<td>3</td>
<td>A single-plant selected from FC 33.243 for its Mi resistance</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Centennial</td>
<td>1977</td>
<td>92</td>
<td>S</td>
<td>1</td>
<td>D64-4636 (Hill×(D49-2491(4)×Jackson)×a tawny pubescent ‘Pickett 71’ type)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Thomas</td>
<td>1989</td>
<td>93</td>
<td>S</td>
<td>1</td>
<td>Centennial × F71-1138(Braxton sib)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Young</td>
<td>1984</td>
<td>96</td>
<td>S</td>
<td>1</td>
<td>Davis × Essex</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sharkey</td>
<td></td>
<td>98</td>
<td>S</td>
<td>1</td>
<td>Tracy × Centennial</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

† R = Ma resistant, MR = moderately resistant to Ma, and S = Ma susceptible.
‡ Allele composition at the L20a locus as seen in Figure 5-1, and at the OPN11 locus in Figure 5-3.
†† The number in parentheses indicates the total number of backcrosses.
In 1896, C.G. Hopkins initiated an experiment to determine the limits of selection for levels of protein and oil in maize. Over the years, the experiment grew more sophisticated as different researchers added nitrogen fertilizers and imposed measures to circumvent selfing (Dudley, 1977). As of 2004, the limits of selection for low oil and low protein had been obtained, but 100 generations of selection had not reached the theoretical ceiling for high oil nor, possibly, for high protein in maize (Dudley and Lambert, 2004). This classic example of plant breeding shows both the promise and exhausting effort of fixing desirable alleles for a trait into an elite group of breeding lines. Similarly to maize, high protein is also a desirable trait in soybean seeds. While soybean oil must compete on the market with other high-quality oils produced by canola (Brassica napus) or sunflower (Helianthus annuus), soybean protein is unique as a high-quality and inexpensive source of vegetable protein. Soybean meal, which is made from soybean seed, delivers approximately 65% of the value of soybeans, and provides 95% of the protein and amino acids incorporated into livestock feed (Chung and Buhr, 1997; Dudley and Lambert, 2004). The USA is the leading producer of soybean meal, exporting it primarily to Canada, Mexico, and the Phillipines, and in total accounting for a value of over $1 billion annually across all foreign markets (American Soybean Association, 2005). Due to its economic importance, a research goal articulated by the United Soybean Board’s Better Bean Initiative (BBI) is to increase the seed protein of U.S. soybean cultivars to better compete with foreign producers that can devote more agricultural land to soybean production (Sallstrom,
If higher protein is provided directly in the seed, the need to concentrate the protein in soybean meal for livestock feed would be eliminated, with substantial savings to the producer (Cober and Voldeng, 2000). In addition to the financial benefits, feeding high-protein soybean to dairy cows gave a 2 kg d\(^{-1}\) boost to milk yield versus that of cows fed with conventional soybean meal (McNiven et al., 1994). Since higher protein in the seeds is sometimes associated with lower levels of the soluble sugars raffinose and stachyose, which are associated with digestive disturbances, high-protein soybean meal could be used at a higher ratio in animal feed without increasing the levels of these sugars (Hartwig et al., 1997). Higher soybean seed protein content is also important in terms of human nutrition, as the U.S. Food and Drug Administration recently approved a recommendation that Americans incorporate as much as 25 g of soy protein into their daily diets to reduce the risk of coronary heart disease, among other health benefits (Hasler, 2002). Both the nutritional value and the low cost of soybean compared to sources of animal protein make it an excellent diet supplement in Africa as well as in Asia (Liu et al., 1995a).

Yet over all the decades soybean has been produced in the USA, the protein content of soybean seeds has held at a steady level of 400 g kg\(^{-1}\) (Wilcox and Guodong, 1997). The constant level of protein is not due to a lack of genetic variation. Soybean germplasm can range in protein content from 347 to 552 g kg\(^{-1}\) (Chung et al., 2003) and numerous studies have reported increased protein content in response to selection. The problem arises instead from the negative correlation of protein with both oil and yield. The strong protein-oil correlation suggests either pleiotropy or a very tight repulsion linkage between protein and oil quantitative trait loci (QTLs), meaning it may be extremely difficult to obtain high-oil lines that also have high protein content. In contrast, the correlation of protein with yield seems less intractable, but
has nevertheless been an obstacle to the development of high-protein, high-yielding lines. Though it is not feasible to conduct over 100 generations of selection to break this negative association of protein with yield, the long-term experiment in maize reveals what is possible given the appropriate tools. This chapter reviews the details of past research with seed protein in soybean, and discusses breeding lines that were obtained in our southern U.S. soybean breeding program that incorporated high protein alleles from an Asian cultivar without a concomitant decrease in yield.

In the late 1960s, soybean oil extracted from the soybean seed was still priced higher than soybean meal. However, as early as 1985 the price of soybean meal was 16% higher than that of soybean oil (Hartwig, 1994). Soybean seeds with lower or higher oil contents are priced accordingly in the market. However, premiums are not yet commonly paid for soybeans that produce soybean meal with significantly higher protein contents because it already has a guaranteed lower limit of protein concentration. Anything above this minimum protein content is considered simply a “gift” to the purchaser. The reason for the lack of premium is due to the difficulty in measuring the protein concentration of a large number of samples, a difficulty that is becoming increasingly resolved by innovations that allow for low-cost, high-throughput sampling. There are also inherent difficulties in preventing the high-protein seeds from mixing with lower-protein seeds at the processor so that growers can be assured of their premium (Greiner, 1990). The lack of a financial incentive, as well as the processing constraints and the negative correlations of seed protein content with both yield and oil, have limited the development of high-protein and high-yielding soybean.

As early as 1932, researchers had noted the negative correlations of protein with other important agronomic qualities such as seed oil and yield, and predicted this would limit the
selection and release of high-protein lines (Weiss et al., 1952). Some researchers have suggested
ignoring the relationship by developing high-protein or high-oil lines for specialty markets
(Brummer et al., 1997). The ideal situation, however, would be to increase both protein and oil,
as well as yield, and the future of high-protein soybean may depend on whether processors will
pay more for these specialty lines or if the negative protein-yield association can be broken
(Leffel and Rhodes, 1993).

Soybean breeders have struggled to develop lines with both high yield and high protein
(Johnson et al., 1955; Shannon et al., 1972). A number of studies have found contradictory
results in the protein-yield association. Though many studies have reported the existence of a
negative correlation of protein with yield, the correlation coefficient was often weak or differed
widely between populations of the same study, leading many authors to conclude that a high
protein and high yielding line was possible (Cober and Voldeng, 2000; Diers et al., 2002;
Hanson et al., 1961; Hartwig and Hinson, 1972; Johnson et al., 1955; Shannon et al., 1972;
Thorne and Fehr, 1970; Xu and Wilcox, 1992). Hartwig (1994) suggested that the negative
correlation reported in many studies may be a combined result of the low-yielding ability of the
high-protein parent and small early-generation populations, rather than a physiological
connection between protein and yield. And indeed he did recover seven F5 lines, each derived
from individual F2 plants, that did not differ significantly in yield from the high-yielding parent,
although they contained as much seed protein content as the high-protein parent (Hartwig, 1994).
The same study also reported success in developing two high-protein lines that were higher
yielding than high oil lines, showing that the high oil-high yield or high protein-low yield
correlation could be overcome (Hartwig and Kilen, 1991). A few papers in the late 1990s
explored the connection between yield and protein as related to plant height, and concluded that
the negative correlation diminished when either determinate lines or the parameter “yield divided by height” were used to calculate the correlation (Mansur et al., 1996; Wilcox and Guodong, 1997). The negative correlation also disappeared with the amount of light intercepted by the soybean plants was increased through thinning or widening the plant rows (Burton, 1994). The majority of studies, however, reported significant negative associations of protein with yield, ranging from as small as –0.07 to as large as –0.62 (Chung et al., 2003; Cober and Voldeng, 2000; Helms and Orf, 1998; Kwon and Torrie, 1964; Orf et al., 1999; Sebolt et al., 2000; Specht et al., 2001; Wehrmann et al., 1987). One study suggested studying correlations only in populations beyond the F₄ generation, since this is the point at which most soybean breeders commence selection; however, even these later generations demonstrated a negative association of –0.2 (Johnson and Bernard, 1962). The weight of evidence thus supports a real negative association of protein and yield, although this relationship may not be significant among certain populations or in cultivars with a determinate growth habit.

The existence of a large negative correlation of protein with oil, however, is more consistent. The correlation coefficient for the relationship of these two traits fell into a wide range from –0.16 to –0.98 (Brim et al., 1959; Chung et al., 2003; Cober and Voldeng, 2000; Fasoula et al., 2004; Hartwig, 1994; Hartwig and Hinson, 1972; Hyten et al., 2004; Johnson and Bernard, 1962; Johnson et al., 1955; Kwon and Torrie, 1964; Lee et al., 1996; Leffel and Rhodes, 1993; Liu et al., 1995b; Mansur et al., 1996; Panthee et al., 2005; Specht et al., 2001; Tajuddin et al., 2003; Thorne and Fehr, 1970; Weiss et al., 1952; Wilcox, 1998; Wilcox and Cavins, 1995; Xu and Wilcox, 1992). Relatively few studies failed to report this negative relationship (Lee et al., 1996; Mansur et al., 1993). The negative correlation between protein and
oil was more pronounced in cultivars that matured later, such as those grown in the southern USA (Yaklich et al., 2002).

The consistent inverse association of these two traits has been expressed as a 2:1 ratio, where the energetic cost of synthesizing two units of protein is approximately equal to the cost of one unit of oil (Hanson et al., 1961). In practice, however, the ratio is often closer to 1.5 to 1.7 (Chung et al., 2003; Hanson et al., 1961; Leffel and Rhodes, 1993), suggesting that protein is either more energetically costly than was originally considered in the equation or that the ratio is genotype-dependent. A key energy cost unique to protein synthesis is the requirement of large amounts of nitrogen. Of 24 vegetable, grain, and legume crops studied for the carbon and nitrogen inputs needed for seed production, soybean had the highest percentage of seed protein, as well as the highest nitrogen requirement (Sinclair and De Wit, 1975).

More effort has been expended to explain the negative protein to oil relationship than the association of protein with yield. Whether the strong association derives from pleiotropy, tight linkage, or the physiological result of both protein and oil drawing from the same limited carbon source remains to be proven. The interrelatedness of these two traits was further confirmed in a study that found that selecting for divergent oil levels in soybean lines inevitably also selected for variation in protein content (Hartwig and Hinson, 1972), a concept also articulated by Hanson et al. (1961). Thorne and Fehr (1970) argued that an increase in protein without a concomitant decrease in oil is only possible if both protein and oil are increased at the expense of lower carbohydrates. This point is underscored by the highly significant negative correlation reported between cell wall polysaccharides in the seed with the combined sum of protein and oil contents, and by the discovery that a genomic region associated with protein and oil contents in multiple studies was also associated with conditioning cell wall polysaccharide content (Stombaugh et al.,
2004). This may suggest that multiple alleles of opposite effects, whether increasing protein, oil, or polysaccharide content, may exist at these loci.

A fascinating paper by Hanson et al. (1961) presciently expands on these points of the protein-oil dichotomy. The authors describe that although genotypes may manifest a “high protein” or “high oil” phenotype or possess varying sizes of carbon reservoirs, the relatively same proportion of carbon across genotypes is available for either protein or oil synthesis. A high-protein phenotype, therefore, is almost always expressed at the expense of oil content, and vice versa. This limitation in the carbon reservoir, also mentioned by Chung et al. (2003), may derive from a lack of genetic variation in the ability to shuttle more carbohydrates into protein or oil synthesis, or simply because a plant can only produce and distribute so much photosynthate to each seed. Higher protein content has been associated with lower carbohydrates, particularly sucrose (Wilcox and Shibles, 2001). Attempting to disrupt photosynthate partitioning may increase protein and oil, but the inevitable decrease in yield would negate these advances.

Though the correlations of protein with yield and oil was accorded more importance in studies over the years, certain researchers also noted associations of protein with other traits and abiotic factors. A pink-flowered soybean mutant was associated with 4 g kg\(^{-1}\) higher protein and 22% higher seed weight (Hegstad et al., 2000a; Stephens et al., 1993). Interestingly, the pink-flowered phenotype appears to be the result of a transposable element called \(Tgm-Express1\), in an allele normally coding for the flavanone 3-hydroxylase gene (Johnson et al., 1998; Zabala and Vodkin, 2005). This locus is putatively involved in both seed protein accumulation and the anthocyanin pathway (Hegstad et al., 2000b). Protein has also been positively correlated with pod dehiscence (Johnson et al., 1955), seed weight (Kane et al., 1997; Panthee et al., 2005), and plant height (Simpson and Wilcox, 1983), and negatively associated with maturity (Diers et al.,
2002; Kwon and Torrie, 1964; Lee et al., 1996; Mansur et al., 1996). Both Simpson and Wilcox (1983) and Xu and Wilcox (1992) reported that lines that matured later tended to have more protein, while an earlier paper reported that the effect of maturity on protein content differed widely between cultivars (Weiss et al., 1952). The contradictory results of the association of two traits, such as maturity and protein content, underline the concept that weak correlations between traits may be confounded with other factors or may be easily broken through genetic recombination in certain soybean populations.

Protein content is also influenced in a positive or negative direction by the environment. Soybean cultivars grown in warmer climates with longer growing seasons tend to have more protein content than those grown in cooler regions, as shown by a comparison of the Northern and Southern Regional Uniform Tests of northern and southern U.S. soybean lines over the past half-century (Yaklich et al., 2002). Another study claimed that lines planted later in the growing season tended to have more protein (Kane et al., 1997). Warm and dry weather has been demonstrated to favor increased protein concentrations, while cooler and wetter weather tends to increase oil concentrations (Dornbos and Mullen, 1992), although the opposite trend has also been reported (Specht et al., 2001).

In economic terms, the negative protein-oil association and the positive yield-oil relationship has meant that high-yielding, high-oil, low-protein cultivars have dominated the soybean market (Leffel, 1989), although this trend seems to be shifting, at least in southern soybean cultivars (Ustun et al., 2001). Of particular note in this debate is the cultivar Essex, which maintains relatively high contents of both oil and protein (Ustun et al., 2001). Other examples bridging this compromise, however, are relatively rare. Earlier studies applied economic theory to this dichotomy, reporting that if a 1% protein increase came at the expense of
a 0.5% decrease in oil or a 0.57% decrease in yield, then that high-protein line would not provide any economic benefit to the industry (Greiner, 1990).

The negative correlation of protein with important agronomic traits, particularly yield, is not the only challenge that soybean breeders attempting to develop higher protein-containing seeds must overcome. Increases in seed protein concentration have been correlated with lower protein quality, particularly in the amino acid balances of lysine, threonine, and the sulfur-containing amino acids (Krishnan et al., 2005; Paek et al., 1997; Panthee et al., 2006; Wilcox and Shibles, 2001; Wilson, 2004). Methionine and cysteine are two sulfur-containing amino acids that are considered essential since humans and monogastric animals cannot synthesize methionine and thus it, along with its downstream product cysteine, must be obtained solely from the diets. Of the two seed protein storage forms, gene expression of the more sulfur-poor protein, 7S β-conglycinin, increases when nitrogen is more abundant, as in high-protein lines (Paek et al., 1997). This results in production of less methionine or cysteine than is required for nutritional purposes, and soybean meal producers must add synthetic amino acids to make up the deficit. There are, however, frequent exceptions to the negative methionine-protein correlation in which either no difference or increased contents of methionine and cysteine were found in high-protein genotypes (Burton et al., 1982; Edwards III et al., 2000; Krober and Cartter, 1966; Wilcox and Shibles, 2001). Additionally, breeding efforts are underway to develop cultivars containing higher levels of these amino acids (Panthee et al., 2004; Panthee et al., 2006). The United Soybean Board’s BBI targets increases in the contents of methionine, cysteine, lysine, and threonine to improve the nutritional value of soybean meal. A favorable amino acid composition of these and other amino acids not only benefits the poultry and livestock feed industry, but may also result in increased gel strength if the concentrations of hydrophobic or
nonpolar amino acids are increased (Riblett et al., 2001; Wilson, 2004). This quality allows processors to more easily use soybean vegetable protein in a range of foods that require greater gel-forming ability, such as tofu or ice cream.

Several studies have examined amino acid differences between cultivars containing either high seed protein content or conventional protein content, although the source of the high-protein genes used in these studies is not explicitly stated. The amino acids associated with nitrogen assimilation, or the amide amino acids, such as asparagine, glutamate, arginine, and glutamine tend to be higher in the high-protein lines (Hernandez-Sebastia et al., 2005; Serretti et al., 1994). Serretti et al. (1994) also reported that tyrosine and alanine increased in the high-protein sources, while threonine, glycine, and histidine tended to decrease. In another study, the amino acids most increased in the high-protein lines, aside from the amide amino acids, were threonine and valine (Hernandez-Sebastia et al., 2005). The effect of protein content on the amount of threonine or valine was mixed in other studies as well; while some reported that both were higher in high-protein lines (Edwards III et al., 2000), others reported that threonine was lower in high-protein lines (Krishnan et al., 2005).

The high heritability estimates of protein content contributed to the success of breeders when selecting for high protein on a phenotypic basis. Heritability was typically calculated on an entry mean basis using variance components, and ranged from 0.2 to 0.96, with an average of 0.63 (Brim and Burton, 1979; Brummer et al., 1997; Chung et al., 2003; Cober and Voldeng, 2000; Diers et al., 1992; Helms and Orf, 1998; Hyten et al., 2004; Kwon and Torrie, 1964; Lark et al., 1994; Lee et al., 1996; Orf et al., 1999; Panthee et al., 2005; Qiu et al., 1999; Sebern and Lambert, 1984; Shannon et al., 1972; Stombaugh et al., 2004; Tajuddin et al., 2003; Wilcox, 1998; Xu and Wilcox, 1992). Encouraged by this, numerous breeding methods were employed
and successfully produced high-protein lines. Results from both single and three-way crosses were reported (Shannon et al., 1972; Thorne and Fehr, 1970). Successful results using the backcross method were also reported; some studies produced high-protein and high-yielding lines after only two backcrosses to the high-yielding parent (Hartwig and Hinson, 1972), while others employed more backcrosses or selfing generations to ensure the high protein alleles were fixed in the progeny (Wehrmann et al., 1987; Wilcox and Cavins, 1995). Interestingly, the latter authors found that each backcross also recovered the higher oil content of the high-yielding parent, suggesting that additional backcrosses may further select against or dilute the effect of the high-protein alleles. Backcrosses became especially advantageous when marker-assisted selection could be used to introgress and maintain high-protein alleles from *G. soja* into agronomically acceptable breeding lines (Sebolt et al., 2000). However, in some cases single crosses were found to be equally effective as backcrosses in recovering high-yielding and high-protein lines (Cober and Voldeng, 2000). Recurrent selection was a choice for a few researchers, and intermated populations were generally advocated to limit the genetic bottleneck that occurs in progeny from single crosses (Hanson et al., 1967). Protein content increased from 2.9 to 8.0 g kg⁻¹ over four to eight cycles of selection, but the method was time-consuming and tedious (Brim and Burton, 1979; Wilcox, 1998; Xu and Wilcox, 1992).

As QTL mapping became more prominent in the early 1990s, it became possible to dissect the protein trait into unique loci that could be manipulated and tested individually in different genetic backgrounds by marker-assisted selection. Each high-protein allele could thus be tested for correlations with yield, oil, seed size, maturity, and amino acid composition, as well as attempting to determine whether the associations were caused by linkage or pleiotropy.
The relative ease of breeding for high protein content and the high heritability for the trait led researchers to estimate that protein content is conditioned by relatively few genes that act additively (Chung et al., 2003; Hanson et al., 1967). QTL mapping has allowed scientists to better understand the genetics of protein content, and has contributed to a number of protein QTL papers in the literature over the past decade (for a review see Chung et al., 2003). In these various studies, QTLs have been detected on almost all linkage groups (LGs), including LG-A1, A2, B1, B2, C1, C2, D1a, D1b, D2, E, F, G, H, I, J, K, L, M, N, O, and P (Brummer et al., 1997; Chapman et al., 2003; Chung et al., 2003; Csanádi et al., 2001; Diers et al., 1992; Hegstad et al., 2000b; Hyten et al., 2004; Lark et al., 1994; Lee et al., 1996; Mansur et al., 1996; Orf et al., 1999; Panthee et al., 2005; Qiu et al., 1999; Sebolt et al., 2000; Specht et al., 2001; Tajuddin et al., 2003; Wang et al., 2002; Weir et al., 2002; Zhang et al., 2004). These results appear to contradict the assessment that relatively few genes control protein content. However, the QTLs on LG-O and P are only reported once in the literature, while the QTLs on LG-A2, B1, B2, C2, D1a, D1b, D2, F, G, H, J, and N are reported two or three times in different papers. QTL validation experiments based on a few previous seed protein mapping studies could confirm on average only 1 of every 2 QTL reported, suggesting that many QTLs may be false positives or are unique only to the population or environment in which they were originally mapped (Fasoula et al., 2004). The QTLs that are detected most consistently, and are thus more likely to represent real QTLs, are the QTLs detected on LG-A1, C1, K and M, which were reported four or five times, the QTLs on LG-E and L, detected six times, and the QTL on LG I, which was detected at least seven times. Hyten et al. (2004) summarizes the map position, LOD score, and the parents used to map these QTLs. The frequency of detection listed here does not mean these QTLs mapped to the same position; that assessment would be difficult to render since even the same
QTL may map to a different position on the linkage group depending on the particular population used.

The QTL detected on LG-I, however, stands out from all other QTL candidates in that it has been detected and mapped to the same position in numerous studies, beginning in the early 1990s (Diers et al., 1992) and summarized over a decade later by Chung et al. (2003). This latter study used composite interval mapping to resolve 11 separate QTLs mapped to similar positions on LG-I into a single large-effect QTL. In many papers, this QTL was mapped using crosses between *G. max* and the wild soybean relative *G. soja*, and was also identified in the Korean accession Pando and the *G. max* accession PI 437088A (Brummer et al., 1997; Chung et al., 2003; Diers et al., 1992; 2002; Sebolt et al., 2000). Danbaekkong, a Korean cultivar developed for its superior protein content (Kim et al., 1996), also contains high-protein alleles at this QTL (Harris, 2001). Danbaekkong’s parents include Dongsan 69, a Korean line, and a high-protein U.S. line, D76-8070, which has a complicated pedigree (Hartwig, 1990), but may have obtained its high-protein allele from ‘Sioux’. Sioux was selected from PI 81021 and donated to the National Plant Germplasm System in 1957 (USDA-ARS/NCGR, 2006). Selective genotyping of high-protein germplasm has confirmed the contribution of the major QTL on LG-I in all 22 different populations (Ritchie et al., 2003). This QTL represents a major locus with potentially multiple alleles. Each high-protein allele at this locus increased protein by 0.89 to 2.50% and decreased oil by 0.45 to 0.75% (Chung et al., 2003; Harris, 2001; Sebolt et al., 2000). The high-protein alleles have also variously been associated with taller plants, smaller seeds, and/or earlier maturity (Chung et al., 2003; Nichols et al., 2006; Sebolt et al., 2000). In relation to this, QTLs affecting maturity and height have been mapped to Satt496 and Satt239, respectively (Abe et al.,
Interestingly, a QTL conditioning high lipid content has also been mapped to this same interval on LG-I (Tajuddin et al., 2003).

Several soybean researchers have since transferred the high-protein alleles at this QTL into adapted lines. Though MAS for the LG-I QTL has been facilitated by the tightly linked SSR markers Satt239 and Satt496, which are separated by 0.5 cM on the genetic map (Song et al., 2004), high-throughput genotyping at this locus in the future may be accomplished with SNP markers (Hwang et al., 2002). Sebolt et al. (2000) reported an approximately 1% increase in protein effected by introgressing the alleles from *G. soja* into three different genetic backgrounds. Diers et al. (2002) backcrossed the alleles into an adapted line multiple times, eventually creating three different BC$_4$ populations from which high-protein and high-yielding lines were identified. By selecting the top 12% of the high-protein lines in a cross of an elite cultivar by the D76-8070 high-protein breeding line, the putative source of the high-protein allele in this study, Hartwig (1994) was able to recover lines that did not significantly differ from the elite parent in yield or from the high-protein parent in protein content. Other studies dealing with populations developed from a high-yielding line crossed to a high-protein line have also reported similar success in recovering single lines within the population that contain both high protein and a yield comparable to the high-yielding parent (Cober and Voldeng, 2000). When Pando, a plant introduction with high-protein alleles at the locus on LG-I, was used as a high-protein donor, approximately 19% of the BC$_2$F$_2$-derived progeny were significantly higher in yield than Pando and significantly higher in protein than the recurrent high-yielding parent (Wehrmann et al., 1987). Wilcox and Cavins (1995) continued breeding with Pando, and were able to recover at least one line that was not significantly different from the high-yielding recurrent parent in yield and contained only 26 g kg$^{-1}$ less protein than Pando. Harris (2001)
used MAS with Satt 239 to screen F2–derived lines from a cross of the highly productive cultivar, Benning, with the high-protein cultivar Danbaekkong. She compared the effect of the Danbaekkong alleles at Satt 239 versus that of the Benning alleles in four environments. Protein was negatively correlated with yield ($r=-0.47^{**}$) over the entire population of lines from this cross. However, a single line, G98SF-114, was homozygous for the Danbaekkong allele at Satt 239, had a protein content of 486 g kg$^{-1}$, and a yield equal to Benning. This line was also homozygous for high-protein alleles at four additional protein QTLs on LG-A1, A2, E, and K, which explained 8 to 11% of the protein variation in the original mapping population. In comparison, the LG-I QTL explained 49% of the variation in this population (Harris, 2001).

**Objectives**

Although G98SF-114 represents a successful introgression of high-protein alleles both at LG-I and on four other LGs, the effect of the Danbaekkong allele at the LG-I QTL on the genetic backgrounds of other elite cultivars is still unknown. While the four minor QTLs on LGs-A1, A2, E, and K do condition some of the increase in protein content and quality, those alleles were not selected during the creation of the backcrossed families, and those QTLs are now fixed for the recurrent parent’s alleles. There are three questions we wish to answer with regard to the high-protein Danbaekkong allele at the QTL on LG-I: i) the extent to which the allele increases protein content in lines derived from G98SF-114 in the genetic backgrounds of three Roundup® cultivars, ii) the effect of this high-protein allele on agronomic qualities, particularly yield, but also seed weight and maturity, among other agronomic traits, and iii) whether there are differences in amino acid profiles, particularly in regards to methionine and cysteine contents. In order to answer such questions, the G98SF-114 high-protein allele at the protein QTL on LG-I was transferred into three elite southern soybean cultivars to compare the effect of the high-
protein allele on yield and other agronomic traits versus that of the elite cultivar’s allele in two years of testing in three locations.

References


CHAPTER 7

AGRONOMIC PERFORMANCE AND AMINO ACID COMPOSITION OF NEAR-ISOGENIC SOYBEAN LINES DIFFERING FOR AN ALLELE AT A MAJOR PROTEIN QTL

Abstract

Although protein provides over half of the economic value of the soybean \textit{[Glycine max \textit{L.} Merr.]} seed, the development of high-protein and high-yielding cultivars is complicated by the negative association of protein content with yield. G98SF-114, a line developed from a cross of the high-protein Korean cultivar Danbaekkong to the cultivar Benning, had a seed protein content of approximately 480 g kg$^{-1}$ and yielded similarly to Benning. From the results of quantitative trait locus (QTL) analysis, G98SF-114 was determined to contain high-protein alleles at a major locus conditioning seed protein content on linkage group (LG) I. The high-protein allele from G98SF-114 was introgressed into three different cultivar backgrounds, Benning-RR, Haskell-RR, and Prichard-RR, with the assistance of the simple sequence repeat (SSR) markers Satt239 and Satt496 that flank the protein locus. Other backcross lines were selected that contained the recurrent parent allele at these markers. The lines were tested in three locations for 2 yr to determine the effect of the presence or absence of the high-protein alleles on agronomic traits and protein and amino acid contents. The high-protein lines consistently matured 4 d earlier, produced seeds that were 7 to 12 mg seed$^{-1}$ smaller, contained 36 to 47 g kg$^{-1}$ more protein and 13 to 22 g kg$^{-1}$ less oil, had higher concentrations of glutamate and other amide amino acids, and lower concentrations of threonine, valine, tyrosine, and lysine residues than lines that did not contain this high-protein allele. The high-protein allele had no significant effect on yield in the three cultivar backgrounds in 2004, but was associated with significantly lower yields in two of the three backgrounds in 2005. However, single lines were recovered that had both high yield and high protein contents in all three backgrounds. The sensitivity of this allele to both genetic background and to different environmental conditions that prevailed in the 2 yr are discussed, as well as potential reasons for its effect on maturity and yield.
Introduction

A research goal articulated by the United Soybean Board’s Better Bean Initiative (BBI) is to increase the seed protein of U.S. soybean cultivars to better compete with foreign producers that can devote more agricultural land to soybean production (Sallstrom, 2002). While soybean oil must compete on the market with other high-quality oils produced by canola (*Brassica napus*) or sunflower (*Helianthus annuus*), soybean protein is unique as an essentially complete and inexpensive source of vegetable protein. Soybean meal, which contains a high percentage of the protein in the seed, accounts for approximately 65% of the value of soybean seed, and provides 95% of the protein and amino acids incorporated into livestock feed (Chung and Buhr, 1997). While the USA is the leading producer of soybean meal, resulting in a value of over $1 billion annually (American Soybean Association, 2005), the protein content and quality of South American soybeans are quickly surpassing those produced in the USA. This is partially a result of increased soybean production in northern U.S. states such as Minnesota, South Dakota, and North Dakota, which tend to produce high-yielding but low-protein soybeans, and decreased production in southern U.S. states, which tend to produce soybeans higher in seed protein content (Yaklich et al., 2002).

There are multiple advantages to further increasing the protein concentration of soybean meal. Even a 1% increase in protein is estimated to add an annual value of $90 to $204 million to the U.S. soybean crop (Chung and Buhr, 1997; Greiner, 1990), partially by reducing the need to concentrate the protein in soybean meal for livestock feed (Cober and Voldeng, 2000). Higher soybean seed protein content is also important for human nutrition, as the U.S. Food and Drug Administration recently approved a recommendation that Americans incorporate as much as 25 g
of soy protein into their daily diets to reduce the risk of coronary heart disease, among other health benefits (Hasler, 2002).

Yet over the period soybean has been extensively produced in the USA, the protein content of soybean seeds has averaged only 400 g kg\(^{-1}\) (Wilcox and Guodong, 1997; Yaklich et al., 2002). The constant level of protein in U.S.-produced soybeans is not due to a lack of genetic variation. Certain accessions in the germplasm contain more than 500 g kg\(^{-1}\) protein (Carter et al., 2004; Chung et al., 2003), and numerous breeders have reported increased protein content in their breeding lines in response to selection (Leffel and Rhodes, 1993; Thorne and Fehr, 1970; Wehrmann et al., 1987; Wilcox and Cavins, 1995). The problem arises from the negative associations of protein with both oil and yield, a factor that has limited the value and development of high-protein soybean cultivars (Leffel and Rhodes, 1993).

The strong negative protein-oil correlation suggests either pleiotropy or a very tight repulsion linkage between protein and oil quantitative trait loci (QTLs). The correlation coefficient for the relationship of these two traits has ranged from \(-0.16\) to \(-0.98\) (Mansur et al., 1996; Specht et al., 2001). Relatively few studies failed to find this negative relationship (Lee et al., 1996; Mansur et al., 1993). In contrast, the negative association of protein with yield seems less intractable, and varied in different studies from \(-0.07\) to \(-0.62\) (Chung et al., 2003; Cober and Voldeng, 2000; Helms and Orf, 1998; Kwon and Torrie, 1964; Orf et al., 1999; Sebolt et al., 2000; Specht et al., 2001; Wehrmann et al., 1987). The association of the two traits, however, was often weak or differed widely between populations of the same study, leading many authors to conclude that the development of a high-protein and high-yielding cultivar was possible (Diers et al., 2002; Hanson et al., 1961; Hartwig and Hinson, 1972; Johnson et al., 1955; Shannon et al., 1972; Thorne and Fehr, 1970; Xu and Wilcox, 1992; Yin and Vyn, 2005).
Seed protein content has also been positively correlated with seed weight (Kane et al., 1997; Panthee et al., 2005) and plant height (Simpson and Wilcox, 1983), and negatively associated with maturity (Diers et al., 2002; Kwon and Torrie, 1964; Lee et al., 1996; Mansur et al., 1996). Both Simpson and Wilcox (1983) and Xu and Wilcox (1992) reported that lines that matured later tended to produce more protein, while an earlier paper reported that the effect of maturity on protein content differed widely between cultivars (Weiss et al., 1952). Increases in seed protein concentration have also been correlated with lower protein quality, particularly in the amino acid balances of lysine, threonine, and the sulfur-containing amino acids (Paek et al., 1997; Panthee et al., 2006; Wilcox and Shibles, 2001; Wilson, 2004). Methionine and cysteine are two sulfur-containing amino acids that are considered essential since humans and monogastric animals cannot synthesize methionine and it along with its downstream product cysteine must therefore be obtained solely from the diet. Any deficiency in the amino acid balance of methionine and cysteine must be supplemented in the diet at additional cost to the soybean processors.

As QTL mapping became more prominent in the early 1990s, it became possible to dissect the protein trait into unique loci that could be manipulated and tested individually in different backgrounds by marker-assisted selection. QTL studies, however, revealed that protein content was a more complex trait than would have been predicted from the relatively high heritability values reported in previous studies. In various experiments, putative QTLs have been detected on almost all linkage groups (LGs), but most often on LGs A1, C1, L, and I (Brummer et al., 1997; Chung et al., 2003; Diers et al., 1992; Hegstad et al., 2000; Lee et al., 1996; Mansur et al., 1996; Orf et al., 1999; Sebolt et al., 2000; Specht et al., 2001; Wang et al., 2002; Weir et al., 2002).
The QTL detected on LG-I, in particular, is unique in that it has been detected and mapped to the same position in numerous studies, and has explained 13 to 80% of the phenotypic variation for protein content in various mapping populations (Brummer et al., 1997; Chung et al., 2003; Diers et al., 1992; Harris, 2001; Sebolt et al., 2000). Selective genotyping of high-protein plant introductions confirmed the contribution of the major QTL on LG-I in all 22 populations evaluated (Ritchie et al., 2003). In several studies, this QTL was mapped using crosses between G. max and the wild soybean relative G. soja (PI 468916), and was also identified in the Korean accession Pando and the G. max accession PI 437088A (Brummer et al., 1997; Chung et al., 2003; Diers et al., 1992; 2002; Sebolt et al., 2000). Danbaekkong, a Korean cultivar developed for its superior protein content (Kim et al., 1996), also contains high-protein alleles at this QTL (Harris, 2001). Danbaekkong’s parents include the Korean cultivar Dongsan 69 and a high-protein U.S. breeding line, D76-8070, which has a complex pedigree (Hartwig, 1990), but may have obtained its high-protein allele from ‘Sioux’, a selection from PI 81021 that was donated to the National Plant Germplasm System in 1957 (USDA-ARS/NCGR, 2006). This QTL thus represents a major locus with potentially multiple alleles.

Marker-assisted selection (MAS) for alleles at the LG-I QTL has been facilitated by the tightly linked simple sequence repeat (SSR) markers Satt239 and Satt496, which flank the locus and are separated by only 0.5 cM on a recently published consensus soybean genetic map (Song et al., 2004). Each high-protein allele at this locus can increase protein content by approximately 6 g kg\(^{-1}\) to 20 g kg\(^{-1}\) (Chung et al., 2003; Harris, 2001; Sebolt et al., 2000), and has variously been associated with taller plants, smaller seeds, and/or earlier maturity (Chung et al., 2003; Nichols et al., 2006; Sebolt et al., 2000).
The high-protein allele(s) on LG-I were also commonly, but not always, associated with lower yield. Diers et al. (2002) backcrossed the high-protein alleles from the *G. soja* accession PI 468916 into an adapted line, eventually creating three different BC4 populations from which high-protein and high-yielding lines were identified. By selecting the top 12% of the high-protein lines in a cross of an elite cultivar by the D76-8070 high-protein breeding line, Hartwig (1994) was able to recover lines that did not differ from the elite parent in yield or from D76-8070 in protein content. Harris (2001) used MAS with Satt239 to screen F2–derived lines from a cross of the elite cultivar Benning with Danbaekkong. There was a negative correlation (r= –0.47**) for protein with yield over the entire population of lines from this cross. However, a single line, G98SF-114, was homozygous for the Danbaekkong allele at Satt239, had a protein content of 48.6%, and a yield equal to Benning. The LG-I QTL at Satt239 explained 49% of the variation in this population (Harris, 2001). This study was initiated to further examine the effect of the high-protein allele from Danbaekkong at the LG-I QTL on agronomic traits and protein quantity and quality in the backgrounds of three different elite cultivars, Benning-RR, Haskell-RR, and Prichard-RR.

**Materials and Methods**

**Plant Material**

The donor of the high-protein allele at the QTL on LG-I, G98SF-114, was crossed to each of the conventional-protein cultivars Benning-RR, Haskell-RR, and Prichard-RR. The F1 plants were crossed again to the elite parents, and this process was repeated for two additional cycles. Since G98SF-114 was selected from a cross between Benning and Danbaekkong, the lines used in this study in the Benning-RR background were BC4-derived, while those in the Haskell-RR and Prichard-RR backgrounds were BC3-derived. In 2003, the backcross-derived F2
plants from the Benning-RR (BC₄) and Haskell-RR (BC₃) backgrounds were genotyped with the simple sequence repeat (SSR) markers Satt239 or Satt496. In the Prichard-RR background, BC₃F₂:₃ plants were genotyped with the same SSR markers. Genomic DNA from each of the backcross-derived F₂ or F₃ plants was extracted from ¼-seed chips cut from the side of each soybean seed opposite the embryo. The chips were placed in individual wells of a 96-well deep-well plate. The seed chips were soaked overnight in 100 µl of sterilized, distilled, and deionized water, and were then macerated with a glass stirring rod and a custom-made crushing device with 96 steel pins. DNA extraction from that point forward followed the protocol in Edwards et al. (1991). The DNA template was diluted to a concentration of approximately 50 ng µL⁻¹ for use in polymerase chain reaction (PCR) using procedures described in Diwan and Cregan (1997). Both Satt239 and Satt496 have a blue (6-FAM) fluorescent tag label (PE-ABI, Foster City, CA). PCR cycling conditions were the same as those used in Diwan and Cregan (1997) with a PE-ABI 9700 thermocycler (PE-ABI, Foster City, CA). The PCR products were separated on a DNA sequencer (ABI PRISM 377, PE-ABI, Foster City, CA). GeneScan® 3.1.2 software (PE-ABI, Foster City, CA) was used to determine the size of the products according to the local Southern option with the ROX 500 ladder (PE-ABI, Foster City, CA) as the size standard.

In the Benning-RR (BC₄) and Haskell-RR (BC₃) backgrounds, 16 backcross-derived F₂ plants were selected for the Benning-RR or Haskell-RR alleles at Satt239 and Satt496, while 16 backcross-derived F₂ plants were selected for the G98SF-114 alleles. The backcross-derived F₂:₃ seeds from each selected plant were sent to the 2004 USDA-ARS winter nursery in Puerto Rico for a seed increase. Eight seeds from each of 14 BC₃F₂:₃ lines of Prichard-RR (4) x G98SF-114 were genotyped to identify within-family F₃ individuals, or pairs, that were homozygous for the parental alleles of either Prichard-RR or G98SF-114 at Satt239 and Satt496. A single pair is
defined as two seeds from the same F2:3 family, with one member of the pair selected for homozygous G98SF-114 alleles at these two SSR markers, and the other member selected for homozygous Prichard-RR alleles with the same markers. Ten pairs were obtained in this manner, and each member of the pair was planted in the greenhouse. The BC3F3:4 lines were grown in the 2004 USDA-ARS winter nursery at Puerto Rico for a seed increase.

The backcross-derived F2:4 lines in the Benning-RR and Haskell-RR backgrounds and the BC3F3:5 lines in the Prichard-RR background were confirmed for the expected SSR genotype at Satt496 using DNA extracted from the leaves. Ten trifoliolate leaves were collected from each line and extracted according to Keim et al. (1988). The DNA was diluted to approximately 50 ng µl⁻¹ and used in a standard PCR (polymerase chain reaction) for microsatellites as described previously. The lines were tested in three field locations in 2004. In 2005, the backcross-derived F2:5 seeds from the Benning-RR and Haskell-RR backgrounds and the BC3F3:6 seeds from the Prichard-RR background that were harvested in 2004 were used to plant three field locations in 2005.

In the 2005 Benning-RR experiment, eight of the entries that had been tested in 2004 were replaced with an additional entry of Benning-RR and with seven BC6F2:4 lines that were homozygous for the G98SF-114 alleles at Satt239. These lines had two additional backcrosses to Benning-RR. No additional lines were added to the Haskell-RR test. In the Prichard-RR background, four entries, or a total of three pairs, that had been tested in 2004 were removed because they were determined to be heterogeneous at Satt239 and Satt496, and the lines were replaced by 13 lines from the BC3F2:4 generation of Prichard-RR (6) x G98SF-114.
Experimental Design

The field tests in 2004 and 2005 were conducted in three environments year\(^{-1}\). Two of the environments were different planting dates at the Univ. of Georgia Plant Sciences Farm near Athens, GA, while the third environment was at the Univ. of Georgia Southwest Research and Education Center at Plains, GA. The two Athens locations differed in planting date by approximately 4 wk in both years. Additional entries within the Benning-RR test in 2004 and the Haskell-RR test in both years included three entries of the recurrent parent and one entry of G98SF-114 within each replication. In 2005, the Benning-RR test contained four entries of Benning-RR and one entry of G98SF-114. In each year, the checks in the Prichard-RR test included five entries of Prichard-RR and two entries of G98SF-114.

The experimental design was a randomized complete block with 36 entries and three replications in both the Benning-RR and Haskell-RR tests. In the Prichard-RR test, there were four replications and 27 entries in 2004, and three replications and 36 entries in 2005. The experimental unit was a two-row plot that was 7-m long with 76-cm between rows, and was seeded with approximately 27 seeds m\(^{-1}\) row. Due to limited seed availability in the 2004 tests, however, the Haskell-RR experiment only had two replications at the Plains location. At maturity, all plots were end-trimmed to a final row length of 3.66 m and the entire plot was harvested with a plot combine.

Data for maturity date, plant height, lodging, seed yield, and seed weight and quality were recorded. Maturity was noted for each plot as the date at which at least 95% of the pods had achieved their mature pod color, referred to as the R8 stage of development (Fehr and Caviness, 1977). Plant height was recorded as the average length of the plant from the ground to the terminal node of three representative plants in the plot. The lodging of the plants in each plot
were given a rating between 1 and 5, where 1 indicates that all plants within the plot are erect and 5 indicates that all plants are prostrate. Both lodging and plant height were noted on mature plants just prior to harvesting. Seed yield is reported on a 130 g kg\(^{-1}\) moisture basis in kg ha\(^{-1}\).

Seed weight was determined from the weight of a sample of 100 seeds from each plot and reported as mg seed\(^{-1}\). Seed quality was judged from 1 (seed surface unblemished with no evidence of shriveling or cracking) to 5 (seeds cracked, shriveled, discolored, or diseased). Upon harvesting the plants, a 25-g seed sample from each plot was sent to the USDA-ARS National Center for Agricultural Research in Peoria, IL to evaluate the protein and oil content (both reported on a moisture-free basis in g kg\(^{-1}\)) using near-infrared transmittance on whole seeds (Hymowitz et al., 1974).

Amino acid profiles were tested on the seeds harvested from three environments in 2004. The amino acid analyses were conducted on 25-g whole seed samples with near-infrared reflectance (NIR). Approximately 6 to 8-g of seeds from each replication of the entries were composited to give the 25-g sample that was submitted for analysis. The samples were analyzed in the laboratory of Dr. Jim Orf (University of Minnesota, St. Paul, MN), in collaboration with Dr. Nick Bajjalieh (Integrative Nutrition Inc., Decatur, IL), and were reported in g kg\(^{-1}\) on a moisture-free basis. Each sample was corrected for its protein content (also in g kg\(^{-1}\)), and the data are therefore expressed in g kg\(^{-1}\) of protein.

Statistical analysis

Agrobase (Generation II, Agronomix Software, Inc., Winnipeg, Manitoba, Canada) was used to conduct analysis of variance for each trait at each location. Replications were considered as random effects and genotypes as a fixed effect in the statistical model. The means across replications for each genotype were combined across environments and years and analyzed by
PROC GLM using SAS (v.8, Cary, NC). For this analysis years and locations were considered random effects. Single degree-of-freedom contrasts were calculated in both the Benning-RR- and Haskell-RR-derived lines to test for differences between the lines containing either the G98SF-114 allele or the Haskell-RR/Benning-RR allele, and between the lines with either allele and the two parents.

Statistical analyses in the Prichard-RR test proceeded essentially as described above, with the following exceptions. Four lines were completely removed from this experiment because it was determined that they had the incorrect genotype and were actually heterogeneous rather than homogeneous for the parental alleles. Single degree-of-freedom contrasts were calculated in the remaining Prichard-RR-derived lines to test for differences between the lines homozygous for the G98SF-114 allele or the Prichard-RR allele, and between the lines and the two parents. Additional contrasts were calculated to compare lines derived from the same family, where one member of the F3 family was homozygous for the G98SF-114 allele at Satt496 and Satt239 and the other was homozygous for the Prichard-RR allele at these markers. There were seven contrasts performed in this manner, and the within-family lines are termed “Prichard pairs”.

Results

Field Experiments Conducted in 2004-05

Weather conditions at the field locations used in 2004 and 2005 differed in both temperature and amount and timing of precipitation during seed filling (Table 7-1). In 2004, over 30 cm of precipitation was received in both Athens and Plains during the beginning of September, while in 2005 only 0 to 1.7 cm of precipitation occurred in the month of September in Athens and Plains, respectively. This corresponds to a critical period of seed filling in the Athens tests, but perhaps not in the Plains tests since these tests tend to mature earlier.
Although the tests were irrigated in 2005, it is likely that the plants still experienced water stress. Temperature also tended to be lower in 2004, averaging 4 to 5ºC less in September than the average temperature in 2005 (Table 7-1).

Based on 2 yr of multiple-environment testing, the Benning-RR-derived BC₄F₂:₄ lines with the high-protein allele from G98SF-114 had approximately 40 g kg⁻¹ greater seed protein content and 17 g kg⁻¹ lower oil content than the BC₄F₂:₄ lines with the conventional protein allele from Benning-RR (Table 7-2). As the seed protein content increased, the yield decreased, with the lines containing the high-protein allele yielding approximately 213 kg ha⁻¹ less than the lines with the conventional-protein allele at the locus on LG-I. However, one line with the high-protein allele, G03MG-97, was not significantly different in yield than Benning-RR and averaged 44 g kg⁻¹ greater protein content than Benning-RR. The high-protein backcross-derived lines had 341 kg ha⁻¹ greater yield and 31 g kg⁻¹ less protein content than the high-protein donor parent, G98SF-114 (Table 7-2). Lines with the high-protein allele in the Benning-RR background also tended to be slightly shorter, have seeds that averaged 12 mg seed⁻¹ less weight, and matured 4 d earlier than backcross-derived lines with the conventional-protein allele. These trends were also seen in the BC₆F₂:₄ lines that were evaluated only in 2005 (Table 7-2). Although based on only one year of testing at three environments, the seven high-protein allele BC₆F₂:₄ lines as a group yielded 336 kg ha⁻¹ less than Benning-RR, matured approximately 5 d earlier, averaged 6 cm shorter and had seeds that were 16 mg seed⁻¹ smaller. Despite the two additional backcrosses, the protein content remained 43 g kg⁻¹ greater in the lines with the high-protein allele compared to Benning-RR. Since each backcross was followed by MAS for the high-protein alleles on LG-I, the protein content is expected to be the result of alleles at that
locus since a high percentage (approximately 99%) of the genome is Benning-RR in the BC6-derived lines.

In general, the Haskell-RR-derived lines showed similar results to that of the Benning-RR-derived lines. While protein content increased by 35 g kg\(^{-1}\) in the lines with the high-protein allele, oil content decreased by 13 g kg\(^{-1}\), and yield by 283 kg ha\(^{-1}\) compared to the lines with the conventional protein allele (Table 7-3). Similarly to the Benning-RR-derived lines, the Haskell-RR-derived lines with the high-protein allele averaged 2 cm shorter, had seeds that were 12 mg seed\(^{-1}\) smaller, and matured 4 d earlier than the Haskell-RR-derived lines with the conventional protein allele (Table 7-3). However, certain lines within the high-protein class were unique in containing both a high seed protein content and a yield comparable to Haskell-RR. G03MG-343, in particular, had a similar yield and 35 g kg\(^{-1}\) more protein compared to Haskell-RR (Table 7-3). G03MG-299 also yielded similarly to Haskell-RR, but contained 43 g kg\(^{-1}\) more protein. Both lines show potential as high-protein and high-yielding cultivars.

The Prichard-RR field test can be discussed on both a protein QTL allele class (Table 7-4) and a BC3-family protein QTL allele class (Table 7-5) basis since pairs of lines were selected for either the high-protein or the conventional protein alleles at Satt496 and Satt239 within F3 families. On a protein QTL allele class basis, the lines did not significantly differ in yield over 2 yr of testing, even though the high-protein lines averaged 47 g kg\(^{-1}\) greater protein content and 22 g kg\(^{-1}\) less oil content than the conventional protein lines (Table 7-4). The high-protein lines in the Prichard-RR background showed a lower protein:oil ratio (approximately 2:1) than the high-protein lines in the Haskell-RR or Benning-RR backgrounds (2.7:1 and 2.4:1, respectively). However, the same trend of earlier maturity, shorter plants, and smaller seeds was observed in the Prichard-RR background for the lines with the high-protein allele (Table 7-4). Lines that
were backcrossed to Prichard-RR two additional times, or BC$_3$F$_2$:4 lines, were included the field tests conducted in 2005, and these yielded 323 kg ha$^{-1}$ lower than Prichard-RR but had 53 g kg$^{-1}$ greater protein content (Table 7-4). The protein content of these BC$_3$F$_2$-derived lines is about 10 g kg$^{-1}$ higher than that of the BC$_3$F$_3$-derived lines (2-yr average) containing the high-protein allele, a likely consequence of including the former lines only in the tests in 2005, a year which was associated with higher protein contents in all of the lines.

On a BC$_3$-family QTL allele class basis, each time the BC$_3$F$_3$-derived line contained the high-protein allele, it tended to mature 2 to 4 d earlier, be 2 to 5 cm shorter at maturity, and produce seeds that were 2 to 13 mg seed$^{-1}$ smaller than its BC$_3$F$_3$-derived sib line with the conventional protein allele (Table 7-5). As in the QTL allele class comparisons, the seed yields among six of the seven pairs did not significantly differ. In the one case in which the high-protein phenotype was associated with lower yield (pair G03MG-876 and G03MG-880), the differences in maturity (4 d), protein content (56 g kg$^{-1}$), and seed size (11 mg seed$^{-1}$) were among the greatest observed between the pairs (Table 7-5). In contrast, G03MG-865 and G03MG-866 were among the pairs that showed the least differences compared to one another, differing in yield by 121 kg ha$^{-1}$, maturity by 2 d, seed size by 2 mg seed$^{-1}$, and protein content by 29 g kg$^{-1}$ (Table 7-5). The variation between sibs of each pair may relate to the extent of recombination in the original F$_2$ plant selected as the progenitor of each F$_3$-derived pair, particularly since the effects were consistent across multiple environments. This would suggest that the high-protein allele acts in a pleiotropic manner but is differentially influenced by the environment, or that it is not pleiotropic and is associated with the agronomic traits through linkage.
Amino Acid Composition

Amino acid profiles differed between high-protein and conventional-protein lines both within each cultivar background and across the three cultivar backgrounds. While the amino acid differences within each cultivar background were in some cases statistically significant (P ≤ 0.05) between the high-protein and conventional-protein lines, the amino acids that were different across all three cultivar backgrounds between the high-protein and conventional-protein lines are of more interest because they signify a general effect of alleles at this locus. Overall, the high-protein allele from G98SF-114 was significantly associated with higher aspartate (1 g kg⁻¹ of protein), glutamate (4.9 g kg⁻¹ of protein), proline (1.3 g kg⁻¹ of protein), and arginine (3.9 g kg⁻¹ of protein) when averaged across the Benning-RR, Haskell-RR, and Prichard-RR backgrounds (Table 7-6). The high-protein allele was significantly associated with lower amounts of threonine (0.7 g kg⁻¹ of protein), valine (0.8 g kg⁻¹ of protein), tyrosine (0.4 g kg⁻¹ of protein), and lysine (0.5 g kg⁻¹ of protein) across the three backgrounds (Table 7-6). The two categories of lines (presence/absence of high-protein allele) did not significantly differ in their contents of serine, glycine, and histidine in any of the three backgrounds. Cysteine concentration did not significantly differ between the two protein QTL allele classes in the Benning-RR or Haskell-RR backgrounds, but was significantly higher (P<0.01) in the high-protein lines for the Prichard-RR background. Methionine content was slightly lower in the high-protein lines in the Benning-RR background, but did not significantly differ between the high-protein or conventional-protein lines in the other two cultivar backgrounds. Across the three cultivar backgrounds, the high-protein allele did not alter the total methionine + cysteine contents versus the conventional protein allele (Table 7-6). The contents of certain carbohydrates were also determined (data not shown). Stachyose content was lower in the high-protein lines in the
Benning-RR and Haskell-RR backgrounds. Sucrose content was also lower in the high-protein lines in the Haskell-RR background, but was unchanged between the protein QTL allele classes of lines in the other two backgrounds.

**Discussion**

In three different backgrounds, the presence of the high-protein allele on LG-I resulted in lines that matured 4 d earlier, produced seeds that were 7 to 12 mg seed\(^{-1}\) smaller, contained 36 to 47 g kg\(^{-1}\) more protein and 13 to 22 g kg\(^{-1}\) less oil, had higher concentrations of glutamate, aspartate, proline, and arginine, and lower concentrations of threonine, valine, tyrosine, and lysine residues than lines that did not contain this high-protein allele. The relative impact of the high-protein allele on these traits varied among the three cultivar backgrounds. The lines in the Haskell-RR and Benning-RR backgrounds tended to respond similarly to the high-protein allele, perhaps not surprisingly since they both belong to maturity group VII (Boerma et al., 1994; Boerma et al., 1997). The cultivar Prichard-RR, in contrast, belongs to MG VIII (Boerma et al., 2001), and lines derived from it responded slightly different in the presence of the high-protein allele. While the high-protein lines in the Haskell-RR and Benning-RR backgrounds averaged 213 to 283 kg ha\(^{-1}\) less yield than their conventional-protein counterparts, the high-protein allele had no negative impact on yield in the Prichard-RR background. The presence of the high-protein allele in the Prichard-RR-derived lines also resulted in a slightly higher protein increase compared to the original cultivar than that observed in the Benning-RR and Haskell-RR backgrounds, an effect confirmed on a within-family basis in the Prichard-RR pairs. The high-protein allele also resulted in significantly higher amounts of cysteine in the Prichard-RR background, but not in the other two cultivar backgrounds, suggesting that the free amino acid concentration or the protein content may also differ within this background. Since Prichard-RR
and its derived lines belong to a later maturity group, the lines are likely to have experienced different environmental conditions at critical stages during vegetative and reproductive development, which can substantially influence seed protein content and yield.

Similarly to the results of this study, high-protein alleles on LG-I from the *G. soja* accession PI 468916 and the *G. max* accession PI 437088A were generally associated with higher protein, lower oil, smaller seeds, earlier maturity, and reduced yield (Chung et al., 2003; Nichols et al., 2006; Sebolt et al., 2000). The effect was consistent enough that QTLs for yield, maturity, and seed size could be mapped to this region of LG-I (Nichols et al., 2006). However, in one of the three populations tested, the high-protein allele did not have a significant effect on seed weight or yield, and two BC$_3$F$_5$-derived lines that arose from the same BC$_3$F$_4$ recombinant plant showed either a negative association between protein with yield and maturity or no association at all (Nichols et al., 2006).

Compared to these results, the high-protein allele in this study seemed to have a greater effect on protein content, with each allele increasing seed protein content by almost 25 g kg$^{-1}$, compared to increases of 6 to 10 g kg$^{-1}$ of seed protein content per allele reported for alleles introgressed from PI 468916 or PI 437088A (Chung et al., 2003; Sebolt et al., 2000). This suggests that the high-protein allele in Danbaekkong may represent a different allele from the high-protein alleles in other sources. The allele in this study also had different effects on agronomic traits than have been reported using other sources of the high-protein allele. In this study, the presence of the high-protein allele was consistently associated with both maturity and seed weight, but was not consistently associated with yield. In the 2004 field tests, there were no significant differences in yield between the high or conventional-protein classes in any of the three genetic backgrounds. However, the yield of the high-protein classes in the Benning-RR
and Haskell-RR backgrounds was significantly lower in 2005, which caused the yield between classes in those two backgrounds to be significantly different across the 2 yr. The inconsistent effect of the high-protein allele in this and other studies could suggest a sensitivity of the trait to different environmental conditions, conditions which occur at different critical time periods in the MG-VII cultivars like Benning-RR and Haskell-RR than in MG-VIII cultivars such as Prichard-RR.

Although the high-protein class in two of the three genetic backgrounds did have lower yield than the conventional protein class, it was possible to find individual lines within the Benning-RR and Haskell-RR backgrounds, G03MG-97 and G03MG-343, respectively, that had high seed protein content but did not differ in seed yield from the recurrent parent. In other studies using populations developed from D76-8070, which contributed the high-protein allele used in this study, and the elite cultivar Forrest, it was also possible to recover high-protein and high-yielding lines (Hartwig, 1994). When Pando was used as a high-protein donor, approximately 19% of the BC₂F₂-derived progeny yielded significantly higher than Pando but had significantly greater seed protein content than the high-yielding recurrent parent (Wehrmann et al., 1987). Wilcox and Cavins (1995) continued breeding with Pando, and recovered at least one line that yielded similarly to the high-yielding parent and contained only 26 g kg⁻¹ less protein than Pando. G98SF-114, the high-protein parent used in this study, was itself only one of six high-protein lines examined that had comparable yield to the high-yielding parent (Harris, 2001).

Protein quality was also an issue addressed in this study. In the presence of the high-protein allele and when evaluated across all three cultivar backgrounds, the soybean seeds contained higher amounts of aspartic acid, glutamic acid, proline, and arginine, lower amounts of
threonine, valine, tyrosine, and lysine, and no alteration in the amounts of serine, glycine, histidine, methionine, and cysteine. This latter result is in accord with other studies that found either no difference or increased contents of methionine and cysteine in a high-protein background (Burton et al., 1982; Edwards III et al., 2000; Krober and Cartter, 1966; Wilcox and Shibles, 2001). Some studies have reported that the amide amino acids, which include asparagine, glutamate, arginine, and glutamine, are associated with nitrogen assimilation in the vegetative and reproductive organs and tend to be higher in high-protein lines (Hernandez-Sebastia et al., 2005; Serretti et al., 1994). Serretti et al. (1994) also reported that tyrosine and alanine increased, while threonine, glycine, and histidine decreased, in high-protein backgrounds. Similarly, Krishnan et al. (2005) reported that threonine content decreased in high-protein lines; however, other studies have reported that threonine content, as well as the content of valine, increased in a high-protein background (Edwards III et al., 2000). Most of the published trends in amino acid differences, such as that of threonine, were reversed or not significantly different in this study, suggesting inconsistent effects of high-protein alleles on non-amide amino acids.

Many of the amino acids mentioned in the previous paragraph are nonessential for human or animal health; however, these amino acid differences alter the amino acid composition of the storage proteins in the seed, which in turn affects industrially important properties such as gel strength and emulsifying ability (Riblett et al., 2001). Previous reports have demonstrated that glutamine promotes the accumulation of β-conglycinin (Bennett et al., 2003; Hernandez-Sebastia et al., 2005), a major storage protein in soybean seeds. Supplemental nitrogen application, which is associated with glutamine accumulation, was also associated with decreased expression of the gene for the Bowman-Birk inhibitor (BBI). The BBI is a minor protein component of soybean
seeds (Krishnan et al., 2005), that, along with the storage protein glycinnin, contains more methionine and cysteine residues than β-conglycinin (Clarke and Wiseman, 2000; Krishnan et al., 2005). β-conglycinin, in contrast, is rich in aspartate, glutamate, leucine, and arginine (Clarke and Wiseman, 2000). Three of these four amino acids were higher in the high-protein lines in this study, possibly suggesting that these lines may accumulate greater amounts of β-conglycinin.

The effect of the high-protein allele on maturity, yield, protein content, and amino acid composition can be considered and explained from a physiological perspective through the self-destruct hypothesis. This hypothesis claims that the nitrogen demand in the seeds can become so great that the fixed nitrogen from the root nodules cannot meet the demand, thereby necessitating nitrogen translocation from the vegetative part of the plant to the seeds (Sinclair and de Wit, 1976). The translocation of nitrogen into the seeds from the leaves occurs in the forms of asparagine and glutamine, which are products from protein degradation in the leaves, and is part of the normal process of senescence within the plant. Indeed, one study calculated that 80% of the soybean plant’s nitrogen was found in the seeds at maturity (Hammond et al., 1951). A high-protein phenotype would presumably exacerbate the self-destructive tendency, possibly causing proteins in the leaves to be degraded at an earlier time point to meet the increased nitrogen demand in the seeds (Saravitz and Raper, 1995; Sinclair and de Wit, 1976). Salado-Navarro et al. (1985) determined that the negative association of high protein with yield in their study was indeed correlated with a greater rate of nitrogen portioning and allocation of dry matter into the seeds, which resulted in a shorter seed filling period and earlier maturity in those lines. Another study reported similar results, but attributed the difference to a stronger observed growth inhibition in the high-protein lines in response to shorter daylengths (Cure et al., 1982).
importance of light quantity and quality was also emphasized by Burton (1994), who maintained high yield in high-protein genotypes by increasing the amount of light. He suggests that the low yield-high protein association is physiological rather than genetic, and is influenced by the supplies of nitrogen and carbon, as well as adequate light conditions, to the plant.

The physiological connection between yield and protein, as well as the sensitivity of the trait to environmental conditions, present obstacles in developing high protein and high yielding lines. In this study, the differences in protein content and maturity between the high-protein and conventional protein classes was much higher in 2005 than in 2004 in all three backgrounds. The different environmental conditions in the 2 yr, particularly in the timing and amount of precipitation, could have influenced the results. There was a 4-wk period from the end of August until the beginning of October in 2005 which the Athens locations did not receive any rainfall. During the same time period, the temperature was warmer than in 2004. These relatively hot and dry conditions have been associated with increased protein concentrations, while cooler and rainy weather tend to increase oil concentrations (Dornbos and Mullen, 1992), although this trend is sometimes reversed (Specht et al., 2001). In 2005, the high-protein lines could have thus accumulated disproportionately more protein as a possible result of both the environmental effect and the genetic effect of the high-protein allele, causing earlier maturity and lower yield than in the lines with the conventional protein allele. As a MG-VIII cultivar, the Prichard-RR lines would have experienced this stress at a different point in their development than in the other two backgrounds, which is a possible reason the yield did not decrease in that test. The sensitivity of this trait both to genetic background and to environmental conditions further complicate attempts to develop high-protein and high yielding lines.
The importance of maintaining a high yield in high-protein lines, which is the ultimate goal in breeding for higher seed protein content, is a topic that has been subject to much debate. Earlier studies indicated that if a 1% protein increase came at the expense of a 0.5% decrease in oil or a 0.57% decrease in yield, then that high-protein line would not provide any economic benefit (Greiner, 1990). Conversely, an economic interpretation suggested that higher-protein lines that yielded less would command a higher market price for the grower than high-protein lines that did not sustain yield losses in that demand for the product would be high but the supply would be low (Chung and Buhr, 1997). Regardless, the economic returns of high-protein soybean will not be fully realized unless the growers are financially rewarded for their high-protein seeds. Although the incorporation of such premiums into the system requires substantial changes, processors such as AGP and Cargill have already initiated premium programs, and seed companies are marketing cultivars that meet the seed protein content requirements to receive the premiums. Such efforts diversify the soybean market, allowing international markets to choose U.S.-produced soybeans with the particular seed traits that best meet their country’s needs.

Acknowledgements

The authors gratefully acknowledge field assistance from Earl Baxter, Gina Rowan, Joe Head, and Herschel Chambers. This work was supported by funding from the American Seed Trade Association, the United Soybean Board, the Georgia Agricultural Experiment Stations, and an NSF pre-doctoral fellowship to J.L. Yates.

References


Agribusiness 13:469-482.

Chung, J., H.L. Babka, G.L. Graef, P.E. Staswick, D.J. Lee, P.B. Cregan, R.C. Shoemaker, and 
J.E. Specht. 2003. The seed protein, oil, and yield QTL on soybean linkage group I. Crop 
Sci. 43:1053-1067.

Clarke, E.J., and J. Wiseman. 2000. Developments in plant breeding for improving nutritional 


Cure, J.D., R. Patterson, C.D. Raper Jr., and W.A. Jackson. 1982. Assimilate distribution in 

Diers, B.W., K. Glover, and S. Carlson. 2002. Fine mapping of a protein QTL on linkage group I 
from Glycine soja. Bienn. Conf. Cell. Mol. Biol. Soybean, 9th, Urbana-Champaign. 11- 

Diers, B.W., P. Keim, W.R. Fehr, and R.C. Shoemaker. 1992. RFLP analysis of soybean seed 

repeat (SSR) markers to assay genetic variation in soybean. Theor. Appl. Genet. 95:723- 
733.


Xu, H., and J.R. Wilcox. 1992. Recurrent selection for maturity and percent seed protein in


from soybean northern and southern region uniform tests. Crop Sci. 42:1504-1515.

Yin, X., and T.J. Vyn. 2005. Relationships of isoflavone, oil, and protein in seed with yield of
Table 7-1. Mean precipitation and maximum and minimum temperatures at Athens and Plains, GA in 2004 and 2005.

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>August</td>
<td>Athens</td>
<td>8.3</td>
<td>12.6</td>
<td>29.8</td>
<td>18.6</td>
<td>20.8</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Plains</td>
<td>13.1</td>
<td>15.6</td>
<td>31.4</td>
<td>20.4</td>
<td>22.0</td>
<td></td>
</tr>
<tr>
<td>September</td>
<td>Athens</td>
<td>33.0</td>
<td>0.0</td>
<td>25.9</td>
<td>17.0</td>
<td>17.9</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Plains</td>
<td>31.2</td>
<td>1.7</td>
<td>27.9</td>
<td>18.9</td>
<td>19.9</td>
<td></td>
</tr>
</tbody>
</table>
Table 7-2. Seed yield, protein and oil contents, and agronomic performance of Benning-RR near-isogenic lines and parents.

<table>
<thead>
<tr>
<th>Environments</th>
<th>Yield</th>
<th>Maturity</th>
<th>Plant height</th>
<th>Plant lodging</th>
<th>Seed weight</th>
<th>Seed quality</th>
<th>Seed Protein</th>
<th>Seed Oil</th>
</tr>
</thead>
<tbody>
<tr>
<td>BC₄F₂:₄:₅</td>
<td>3175</td>
<td>10/20</td>
<td>101</td>
<td>2.8</td>
<td>134</td>
<td>1.8</td>
<td>401</td>
<td>198</td>
</tr>
<tr>
<td>Ben/Ben</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Benning-RR</td>
<td>3193</td>
<td>10/18</td>
<td>102</td>
<td>2.9</td>
<td>132</td>
<td>1.7</td>
<td>396</td>
<td>202</td>
</tr>
<tr>
<td>G98SF-114</td>
<td>2620</td>
<td>10/19</td>
<td>81</td>
<td>2.2</td>
<td>147</td>
<td>1.6</td>
<td>475</td>
<td>182</td>
</tr>
</tbody>
</table>

† Values followed by the same letter/number do not significantly differ at the p=0.05 level.

‡ Lodging score where 1 indicates all plants in the plot are erect, and 5 indicates all plants in the plot are prostrate.

§ Seed quality score where 1 indicates seeds perfect and 5 indicates seeds are cracked, discolored, diseased, etc.

¶ Ben/Ben = homozygous for the Benning-RR allele at the LG-I protein QTL.

†† 114/114 = homozygous for the G98SF-114 allele at the LG-I protein QTL.
Table 7-3. Seed yield, protein and oil contents, and agronomic performance of Haskell-RR near-isogenic lines and parents.

<table>
<thead>
<tr>
<th></th>
<th>Yield</th>
<th>Maturity</th>
<th>Plant height</th>
<th>Plant lodging</th>
<th>Seed weight</th>
<th>Seed quality</th>
<th>Seed protein</th>
<th>Seed oil</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>kg ha(^{-1})</td>
<td>date</td>
<td>cm</td>
<td>score(^{\ddagger})</td>
<td>mg seed(^{-1})</td>
<td>score(^{\S})</td>
<td>---g kg(^{-1})---</td>
<td></td>
</tr>
<tr>
<td>BC(_3)F(_2)-4-5 (Has/Has)(^{\mathcal{P}})</td>
<td>3323 (^{\ddagger})</td>
<td>10/21 (^{b})</td>
<td>100 (^{a})</td>
<td>2.7 (^{b})</td>
<td>147 (^{a})</td>
<td>1.8 (^{a})</td>
<td>397 (^{c})</td>
<td>199 (^{a})</td>
</tr>
<tr>
<td>BC(_3)F(_2)-4-5 (114/114)(^{\ddagger})</td>
<td>3040 (^{b})</td>
<td>10/17 (^{c})</td>
<td>98 (^{b})</td>
<td>2.9 (^{a})</td>
<td>135 (^{b})</td>
<td>1.8 (^{a})</td>
<td>432 (^{b})</td>
<td>186 (^{b})</td>
</tr>
<tr>
<td>G03MG-343 (114/114)</td>
<td>3237</td>
<td>10/18</td>
<td>101</td>
<td>2.7</td>
<td>143</td>
<td>1.8</td>
<td>427</td>
<td>188</td>
</tr>
<tr>
<td>G03MG-299 (114/114)</td>
<td>3173</td>
<td>10/20</td>
<td>100</td>
<td>2.8</td>
<td>137</td>
<td>1.8</td>
<td>435</td>
<td>184</td>
</tr>
<tr>
<td>Haskell-RR</td>
<td>3208 (^{a})</td>
<td>10/22 (^{a})</td>
<td>100 (^{a})</td>
<td>2.7 (^{ab})</td>
<td>148 (^{a})</td>
<td>1.9 (^{a})</td>
<td>392 (^{c})</td>
<td>200 (^{a})</td>
</tr>
<tr>
<td>G98SF-114</td>
<td>2700 (^{c})</td>
<td>10/17 (^{c})</td>
<td>88 (^{c})</td>
<td>2.5 (^{b})</td>
<td>137 (^{b})</td>
<td>1.9 (^{a})</td>
<td>459 (^{a})</td>
<td>176 (^{c})</td>
</tr>
</tbody>
</table>

\(^{\ddagger}\) Values followed by the same letter/number do not significantly differ at the p=0.05 level.

\(^{\ddagger}\) Lodging score where 1 indicates all plants in the plot are erect, and 5 indicates all plants in the plot are prostrate.

\(^{\S}\) Seed quality score where 1 indicates seeds perfect and 5 indicates seeds are cracked, discolored, diseased, etc.

\(^{\mathcal{P}}\) Has/Has = homozygous for the Haskell-RR allele at the LG-I protein QTL.

\(^{\ddagger}\) Has/Has = homozygous for the G98SF-114 allele at the LG-I protein QTL.
Table 7-4. Seed yield, protein and oil contents, and agronomic performance of Prichard-RR near-isogenic lines and parents.

<table>
<thead>
<tr>
<th>Environments</th>
<th>Yield no. kg ha⁻¹</th>
<th>Maturity date</th>
<th>Plant height cm</th>
<th>Plant lodging score ‡</th>
<th>Seed weight mg seed⁻¹</th>
<th>Seed quality score §</th>
<th>Seed protein g kg⁻¹</th>
<th>Seed oil g kg⁻¹</th>
</tr>
</thead>
<tbody>
<tr>
<td>BC₃F₃:4‑5 (Pri/Pri)</td>
<td>6</td>
<td>2620b†</td>
<td>10/31ᵃ</td>
<td>108ᵃ</td>
<td>3.4ᵃ</td>
<td>126ᵃ</td>
<td>1.9ᵃ</td>
<td>414ᵃ</td>
</tr>
<tr>
<td>BC₃F₃:4‑5 (114/114)††</td>
<td>6</td>
<td>2600ᵇ</td>
<td>10/27ᵇ</td>
<td>104ᵇ</td>
<td>3.3ᵃ</td>
<td>119ᵇ</td>
<td>1.8ᵇ</td>
<td>461ᵇ</td>
</tr>
<tr>
<td>Prichard (RR)</td>
<td>6</td>
<td>2755ᵃ</td>
<td>10/30ᵃ</td>
<td>104ᵇ</td>
<td>3.3ᵃ</td>
<td>129ᵇ</td>
<td>1.8ᵃᵇ</td>
<td>408ᶜ</td>
</tr>
<tr>
<td>G98SF‑114</td>
<td>6</td>
<td>2520ᵇ</td>
<td>10/19ᵉ</td>
<td>90ᵉ</td>
<td>2.5ᵃ</td>
<td>138ᵃ</td>
<td>1.8ᵇ</td>
<td>173ᶜ</td>
</tr>
<tr>
<td>BC₅F₂:4 (114/114)</td>
<td>3</td>
<td>2788²</td>
<td>10/27²</td>
<td>99¹</td>
<td>2.4¹</td>
<td>127³</td>
<td>1.7¹</td>
<td>471²</td>
</tr>
<tr>
<td>Prichard</td>
<td>3</td>
<td>3111¹</td>
<td>10/30¹</td>
<td>99¹</td>
<td>2.4¹</td>
<td>142²</td>
<td>1.8¹</td>
<td>418³</td>
</tr>
<tr>
<td>G98SF‑114</td>
<td>3</td>
<td>2909¹,²</td>
<td>48³</td>
<td>89²</td>
<td>1.9²</td>
<td>146¹</td>
<td>1.7¹</td>
<td>484¹</td>
</tr>
</tbody>
</table>

† Values followed by the same letter/number do not significantly differ at the p=0.05 level.

‡ Lodging score where 1 indicates all plants in the plot are erect, and 5 indicates all plants in the plot are prostrate.

§ Seed quality score where 1 indicates seeds perfect and 5 indicates seeds are cracked, discolored, diseased, etc.


† Pri/Pri = homozygous for the Prichard-RR allele at the LG-I protein QTL.

†† 114/114 = homozygous for the G98SF-114 allele at the LG-I protein QTL.
Table 7-5. Seed yield, protein and oil contents, and agronomic performance of Prichard-RR sib pairs with or without the high-protein allele.

<table>
<thead>
<tr>
<th>Pair</th>
<th>Line</th>
<th>Yield</th>
<th>Maturity</th>
<th>Plant height</th>
<th>Plant lodging</th>
<th>Seed weight</th>
<th>Seed quality</th>
<th>Seed protein</th>
<th>Seed oil</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>MG03-803 (Pri/Pri)</td>
<td>2614a</td>
<td>61a</td>
<td>107a</td>
<td>3.1a</td>
<td>131a</td>
<td>1.8a</td>
<td>412b</td>
<td>198a</td>
</tr>
<tr>
<td></td>
<td>MG03-805 (114/114)††</td>
<td>2499a</td>
<td>58b</td>
<td>106b</td>
<td>3.4a</td>
<td>126b</td>
<td>1.8a</td>
<td>467a</td>
<td>176b</td>
</tr>
<tr>
<td>2</td>
<td>MG03-810 (Pri/Pri)</td>
<td>2681a</td>
<td>61a</td>
<td>110a</td>
<td>3.4a</td>
<td>128a</td>
<td>1.8a</td>
<td>410b</td>
<td>199a</td>
</tr>
<tr>
<td></td>
<td>MG03-813 (114/114)††</td>
<td>2862a</td>
<td>57b</td>
<td>106b</td>
<td>3.4a</td>
<td>115b</td>
<td>1.8a</td>
<td>464a</td>
<td>176b</td>
</tr>
<tr>
<td>3</td>
<td>MG03-830 (Pri/Pri)</td>
<td>2533a</td>
<td>61a</td>
<td>106a</td>
<td>3.1a</td>
<td>128a</td>
<td>1.9a</td>
<td>410b</td>
<td>197a</td>
</tr>
<tr>
<td></td>
<td>MG03-825 (114/114)</td>
<td>2641a</td>
<td>58b</td>
<td>104b</td>
<td>3.2a</td>
<td>126b</td>
<td>1.8b</td>
<td>460b</td>
<td>176b</td>
</tr>
<tr>
<td>4</td>
<td>MG03-838 (Pri/Pri)</td>
<td>2580a</td>
<td>60a</td>
<td>107a</td>
<td>3.5a</td>
<td>131a</td>
<td>1.9a</td>
<td>412b</td>
<td>197a</td>
</tr>
<tr>
<td></td>
<td>MG03-835 (114/114)</td>
<td>2654a</td>
<td>58b</td>
<td>102b</td>
<td>3.3a</td>
<td>121b</td>
<td>1.7b</td>
<td>460b</td>
<td>178b</td>
</tr>
<tr>
<td>5</td>
<td>MG03-865 (Pri/Pri)</td>
<td>2647a</td>
<td>60a</td>
<td>107a</td>
<td>3.5a</td>
<td>124a</td>
<td>1.9a</td>
<td>424b</td>
<td>196a</td>
</tr>
<tr>
<td></td>
<td>MG03-866 (114/114)</td>
<td>2526a</td>
<td>58b</td>
<td>103b</td>
<td>3.2a</td>
<td>122a</td>
<td>1.8b</td>
<td>453b</td>
<td>176b</td>
</tr>
<tr>
<td>6</td>
<td>MG03-875 (Pri/Pri)</td>
<td>2620a</td>
<td>60a</td>
<td>111a</td>
<td>3.4a</td>
<td>120a</td>
<td>1.8a</td>
<td>417b</td>
<td>199a</td>
</tr>
<tr>
<td></td>
<td>MG03-877 (114/114)</td>
<td>2654a</td>
<td>56b</td>
<td>106b</td>
<td>3.4a</td>
<td>113b</td>
<td>1.8a</td>
<td>459b</td>
<td>176b</td>
</tr>
<tr>
<td>7</td>
<td>MG03-876 (Pri/Pri)</td>
<td>2688a</td>
<td>60a</td>
<td>112a</td>
<td>3.5a</td>
<td>121a</td>
<td>1.8a</td>
<td>410b</td>
<td>199a</td>
</tr>
<tr>
<td></td>
<td>MG03-880 (114/114)</td>
<td>2399b</td>
<td>56b</td>
<td>108b</td>
<td>3.3a</td>
<td>110b</td>
<td>1.8a</td>
<td>466b</td>
<td>176b</td>
</tr>
</tbody>
</table>

† Values followed by the same letter/number do not significantly differ at the p=0.05 level.

‡ Lodging score where 1 indicates all plants in the plot are erect, and 5 indicates all plants in the plot are prostrate.

§ Seed quality score where 1 indicates perfect seeds and 5 that seeds are discolored, diseased, etc.

¶ Pri/Pri = homozygous for the Prichard-RR allele at the LG-I protein QTL.

†† 114/114 = homozygous for the G98SF-114 allele at the LG-I protein QTL.
Table 7-6. Mean amino acid contents between lines containing the high-protein allele from G98SF-114 or the conventional-protein allele averaged across near-isogenic lines from Benning-RR, Haskell-RR, or Prichard-RR.

<table>
<thead>
<tr>
<th>Amino acid</th>
<th>Amino acid content</th>
<th>High-protein lines</th>
<th>Conventional-protein lines</th>
<th>Recurrent parent</th>
<th>G98SF-114</th>
</tr>
</thead>
<tbody>
<tr>
<td>Aspartate</td>
<td></td>
<td>110.9</td>
<td>109.9**</td>
<td>109.8*</td>
<td>111.5</td>
</tr>
<tr>
<td>Glutamate</td>
<td></td>
<td>177.9</td>
<td>173.0***</td>
<td>172.7***</td>
<td>180.8*</td>
</tr>
<tr>
<td>Proline</td>
<td></td>
<td>51.0</td>
<td>49.7***</td>
<td>49.4***</td>
<td>51.7*</td>
</tr>
<tr>
<td>Arginine</td>
<td></td>
<td>74.9</td>
<td>71.0***</td>
<td>71.0***</td>
<td>77.9**</td>
</tr>
<tr>
<td>Serine</td>
<td></td>
<td>43.8</td>
<td>43.7</td>
<td>43.8</td>
<td>43.1</td>
</tr>
<tr>
<td>Glycine</td>
<td></td>
<td>41.7</td>
<td>41.9</td>
<td>41.9</td>
<td>41.4</td>
</tr>
<tr>
<td>Histidine</td>
<td></td>
<td>25.9</td>
<td>25.9</td>
<td>25.9</td>
<td>26.0</td>
</tr>
<tr>
<td>Cysteine</td>
<td></td>
<td>14.5</td>
<td>14.4</td>
<td>14.5</td>
<td>13.8*</td>
</tr>
<tr>
<td>Methionine</td>
<td></td>
<td>14.2</td>
<td>14.3</td>
<td>14.4</td>
<td>13.7</td>
</tr>
<tr>
<td>Threonine</td>
<td></td>
<td>36.2</td>
<td>36.9***</td>
<td>36.8**</td>
<td>35.3**</td>
</tr>
<tr>
<td>Valine</td>
<td></td>
<td>47.6</td>
<td>48.4**</td>
<td>48.7**</td>
<td>46.4*</td>
</tr>
<tr>
<td>Tyrosine</td>
<td></td>
<td>34.3</td>
<td>34.7***</td>
<td>34.6**</td>
<td>33.8*</td>
</tr>
<tr>
<td>Lysine</td>
<td></td>
<td>63.2</td>
<td>63.7**</td>
<td>63.9**</td>
<td>62.3*</td>
</tr>
</tbody>
</table>

* Significant at the 0.05 probability level.

** Significant at the 0.01 probability level.

*** Significant at the 0.001 probability level.
CHAPTER 8

SUMMARY

Soybean is an increasingly important crop in the USA. The protein and oil components of the soybean seeds have almost limitless uses in food products as well as in industrial uses such as paints, inks, diesel fuel, and lubricants. Soybean meal is invaluable to the poultry and livestock industries, and is also increasingly important in aquaculture and as a component of pet foods. In terms of human nutrition, soybean has been shown to treat or help prevent health problems varying from heart disease to cancer. As soybean’s health benefits have become more widely known, the number of soy-based food products has literally exploded on grocery shelves. More people seem to associate soybean with healthy eating and meat alternatives than with trans-fats, hydrogenation, or vegetable oil. Soybean may be elevated in terms of the public perception, but the issues of trans-fats, new diseases, and increasing competition from South American countries are important to U.S. soybean breeders, producers, and processors. Soybean breeders are working to improve the fatty acid profile of soybean seeds, an effort intended to increase the health benefits of the oil and to extend the shelf life of soybean-containing products without using hydrogenation, which creates trans-fatty acids. New diseases such as Asian soybean rust have entered the USA just in the past year, and pose a massive threat to soybean production. The hectareage devoted to soybean production in Brazil is steadily increasing, and the higher-protein soybean seeds exported out of this South American country are preferred in the large Asian markets of China, Korea, and Japan. A plant breeder wishing to work on soybean
has numerous breeding objectives from which to choose, and many priorities to balance in a cultivar development program.

Complicating the issue is the limited diversity that is present in U.S. soybean cultivars. Soybean breeders increasingly rely on the USDA Soybean Germplasm Collection, which contains diverse accessions of both *G. max* and its wild relative *G. soja*, for sources of novel genes providing resistance to pests and pathogens, or for improving agronomic qualities such as protein content and yield. Before the novel genes can be employed in elite cultivars, however, it is necessary to fully characterize and test the effect of the genes in elite cultivar backgrounds. This process involves the steps of identifying the germplasm sources containing the desirable genes, confirming that these genes are unique from previously identified genes that have already been characterized and introduced into cultivars, mapping the position and determining the effect of the gene on the phenotype of interest, and finally, determining the influence of the gene on yield and other important agronomic traits in elite soybean cultivars. The dissertation research presented here considered all of these steps in two different projects. In this study, germplasm sources and cultivars from Japan, China, and Korea were studied to 1) identify and map QTLs conditioning Ma resistance in PIs from China for which there was evidence of novel Ma resistance genes, 2) determine whether the source of Ma resistance in the Chinese PIs and the Japanese PI 200538 was already present in the southern U.S. soybean germplasm, and 3) test the effect of a high-protein allele from a Korean cultivar on important agronomic traits such as protein, yield, maturity, and seed size.

Peanut root-knot nematodes (RKN) are one of four economically important RKN species that altogether cause more global damage to the food supply than any other category of plant-
parasitic nematodes. Peanut RKN (*Meloidogyne arenaria* or Ma) is an increasingly common pest in the southern region in which crops such as peanut and soybean are grown. The current level of Ma resistance in elite southern soybean cultivars is not adequate to prevent yield loss in heavily infested fields; however, this species is increasing in population density in the southeastern USA because of its ability to parasitize both soybean [*Glycine max* (L.) Merr.] and peanut, a crop of historical importance in Georgia, and because populations of other RKN species have traditionally been targeted. This project had three major components: i) to characterize newly identified Chinese sources of Ma resistance for the presence of resistance genes that were unique from those previously identified in PI 200538, ii) to identify the location and effect of the resistance genes in the PIs that were identified in the first objective as containing new sources of Ma resistance genes, and iii), to characterize the Ma resistance in both historically and currently grown soybean cultivars to determine whether the source of resistance in PI 200538 and the Chinese PIs was also present within the southern U.S. soybean germplasm.

In the first objective, a total of four plant introductions (PIs) that had been previously identified as highly Ma resistant were all crossed to the only characterized source of Ma resistance, PI 200538. Populations developed from PI 594403, PI 594427C, and PI 594651L2, each crossed to PI 200538, were evaluated for Ma response in the greenhouse. Progeny from PI 200538 x PI 230977 were also included to confirm previously published results on the uniqueness of genes contained in these two sources. While Ma resistant, PI 230977 had previously been characterized only for genes conditioning resistance to the Javanese RKN (*Mj, M. javanica*). The presence of both highly galled and highly resistant F2 plants was observed in each of the four crosses, and provided evidence that unique resistance genes were segregating
among each population of F₂ individuals. F₂ plants with extreme responses to Ma were transplanted to obtain F₂-derived progeny for two confirmation experiments. In the crosses PI 200538 x PI 594403, PI 200538 x PI 594427C, and PI 200538 x PI 594651L2, the highly galled F₂ plants had F₂:3 progeny that sustained significantly greater galling (P<0.001) in both confirmation experiments than the F₂:3 progeny derived from highly resistant F₂ plants. While the F₂-derived progeny from PI 200538 x PI 594403 were more variable within each category than was observed in the other crosses, the means of the highly galled and highly resistant lines were significantly different. Only the F₂-derived progeny of PI 200538 x PI 230977 failed to perform similarly to the original F₂ plants (P>0.4 in both confirmation studies). However, single lines within this population, as well as in the other populations, were recovered that had fewer or more galls compared to one another or compared to the parents, indicated that resistance genes were segregating among the lines. This was observed even within the PI 200538 x PI 230977 population, suggesting that the minor QTL in PI 230977 that was originally reported to condition Mj resistance may also provide protection against Ma as well. Lines with significantly fewer galls than PI 200538 were not recovered in the PI 200538 x PI 594403 population, suggesting that this source may not contain any large-effect resistance alleles. Alternatively, its variable performance in the greenhouse may indicate that it is heterogeneous for Ma resistance, and may need to be purified before Ma resistance can be accurately characterized.

Since this first objective provided evidence that PI 594427C, PI 594651L2, and possibly PI 594403 were candidates for providing new sources of Ma resistance, the second objective involved identifying the QTLs responsible for the Ma resistance and quantifying the effect of these QTLs. In two of the three soybean plant introductions (PIs), PI 594427C and PI
594651L2, a major QTL on LG-F was discovered that accounted for as much as 41% and 17% of the Ma galling response in two different populations (F2 and F2:3) created from each of the PIs crossed to the susceptible cultivar CNS. In both generations of each cross, the same QTL was identified and thus confirmed, each time mapping to approximately the same position on LG-F, showing a tight association between the PCR (polymerase chain reaction)-based markers L20a or OPN11, and accounting for a similar amount of phenotypic variation. In addition, a minor QTL on LG-E that accounted for as much as 6% of the variation was mapped and confirmed in PI 594651L2. Although minor QTLs were discovered in the F2 population of PI 594403 x CNS, none of them were confirmed in the F2:3 generation; therefore, this PI contains as of yet undetected resistance genes. Evidence from the first study, however, also may indicate that this seed source is heterogeneous for Ma resistance, and that it may not be possible to identify its putative resistance alleles until the seed source is purified.

Both the major QTL on LG-F and the minor QTL on LG-E reported in this study have been previously associated with Ma resistance in the germplasm source PI 200538. The QTL on LG-F in that study was mapped closest to OPN11. The interval between OPN11 and L20a was estimated to be approximately 2 cM in the F2 population of PI 594651L2 x CNS, meaning that QTLs that are more closely associated with L20a or OPN11 in different mapping populations are likely to represent resistance alleles at the same locus. These results confirm the importance of this locus on LG-F, which is positioned within a cluster of genes conditioning resistance to multiple pests and pathogens, and establish the necessity for creating more high-throughput markers, possibly using L20a as first candidate, to allow soybean breeders to more easily introgress this source of Ma resistance into their cultivars.
Since the information from the second objective provided evidence that the Ma resistance allele on LG-F was a major source of Ma resistance, a pedigree analysis study was undertaken to determine whether this unique and effective source of Ma resistance was already present in Ma resistant soybean cultivars. A panel of 61 lines, plant introductions (PIs) and cultivars were therefore evaluated for Ma response as well as allele composition at L20a and OPN11. These two particular markers were chosen because of their close association with the major QTL on LG-F in all three sources of Ma resistance, PI 200538, PI 594427C, and PI 594651L2. The allelic variation in OPN11 did not perfectly correlate with the Ma response, and instead showed evidence in at least one lineage of being in linkage repulsion phase to the resistance gene. In addition, Ma resistant PIs contained different OPN11 alleles. While this could suggest different resistance alleles that belong to the same locus, the observation that ‘CNS’, which is Ma susceptible, and PI 594427C shared the same allele indicated that this locus was not directly responsible for Ma resistance. Variation at OPN11, however, was informative in determining that ‘Palmetto’ was the source of Ma resistance in many cultivars commonly used as parents in southern soybean breeding programs, such as ‘Bragg’, ‘Forrest’, and ‘Braxton’.

In contrast, the allelic variation at L20a perfectly corresponded to the published Ma response in all 61 soybean PIs, lines, and cultivars. PI 594427C, PI 594651L2, and PI 200538 all contained the L20a allele associated with Ma resistance, further strengthening the association of L20a allelic variation to Ma response. Since L20a corresponds to a resistance gene candidate, it is possible that this gene is either tightly linked to the Ma resistance gene or is directly providing the Ma resistance itself. Regardless, L20a allelic variation is highly predictive of Ma resistance and could be effectively used as part of a marker-assisted selection strategy to incorporate Ma
resistance into soybean breeding lines. Although the U.S. southern soybean germplasm already contains at least one allele of the Ma resistance gene on LG-F, the three Ma-resistant PIs could contain unique resistance alleles of this gene that could provide a higher level of Ma resistance. Further study of the allele sources from the PIs versus those introduced by Palmetto and S-100 is warranted to fully incorporate the alleles that provide the highest level of Ma resistance to southern soybean cultivars.

Thus far, the steps towards identifying and characterizing germplasm sources have focused solely on peanut root-knot nematode resistance. If this project had been continued, one of the next steps would have included evaluating the effect of the Ma resistance allele on agronomic traits such as yield. This part of the process, however, was considered with a different project, which determines the effect of a high-protein allele on yield, protein and amino acid contents, as well as other agronomic qualities. The high-protein allele in the Korean cultivar Danbaekkong was transferred into three southern elite U.S. soybean cultivars using DNA markers that flank the gene(s) conferring the high seed protein content. In multiple environment field tests comparing the effect of the presence or absence of the high-protein allele in the three cultivar backgrounds on such agronomic qualities as yield, protein content and quality, maturity, and seed weight, it was determined that the high-protein allele was consistently associated with seed protein content that was 36 to 47 g kg\(^{-1}\) higher than when the high-protein allele was absent, oil content that was 13 to 22 g kg\(^{-1}\) lower, plants that matured approximately 4 d earlier, seeds that were 7 to 12 mg seed\(^{-1}\) smaller, and higher concentrations of amino acids associated with nitrogen assimilation. Its effect on yield, which is the most important trait to soybean producers, was inconsistent and found to depend on both the genetic background and on
the environmental conditions. While the high-protein allele had no effect on yield in the Prichard-RR genetic background, it was associated with lower yield in 2005, but not in 2004, in the Benning-RR and Haskell-RR backgrounds. Several factors differed between the two yr. While 2005 experienced a great amount of rain throughout the growing season except for the entire month of September in the Athens locations, the growing season in 2004 experienced a great amount of precipitation in early September from two different hurricanes. As a MG VIII cultivar, the Prichard-RR background experienced the timing of the precipitation at different points in its vegetative and reproductive development than did the earlier maturing cultivars Haskell-RR and Benning-RR (both MG VII). Since 2005 was generally a warmer year than 2004, particularly during September, this, along with the lack of precipitation, could have differentially influenced the protein contents and yield of the near-isogenic lines. These and other unknown factors resulted in lines containing the high-protein allele maturing 1 to 2 d earlier and containing at least 10 g kg\(^{-1}\) more seed protein content than was observed in the same lines that were grown in 2004.

The effect of the high-protein allele on yield, maturity, seed weight, and amino acid and protein contents may be a result of a phenomenon described as the self-destruct hypothesis. This trend, which involves protein degradation and ultimately senescence in the leaves to meet the high nitrogen demand in the growing seeds, is a condition that would be presumably intensified in high-protein lines. The hypothesis explains why glutamate and other amino acids involved in the nitrogen metabolism and remobilization in the plant would be found at a higher concentration in the seeds of high-protein lines. The increased allocation of nitrogen in the form of glutamine or asparagine from the leaves to the seeds would result in faster degradation of proteins and thus
earlier senescence of the leaves. Ultimately, the plant would mature earlier, and may have smaller seeds and reduced yields as a result of a shorter seed filling period compared to a line in which the nitrogen demand in the seeds is not as strong. Under such a scenario, one would expect the yield of high-protein lines to be consistently lower than that of conventional protein lines. Instead, most studies have been able to recover a few lines from their high protein x high yielding crosses that retained desirable qualities from both parents. Even in the Haskell-RR and Benning-RR backgrounds in 2004 and 2005, it was possible to find single lines that had both high protein and high yield. It is not known how or why some lines can overcome the negative yield-protein association and maintain both high yield and high seed protein content in multiple environments, while others are constrained by the negative protein-yield association in some environments, but not in others. Perhaps the correlation between these two traits has both a genetic and physiological basis, and even if the genetic linkage between yield and protein is broken, the physiological correlation still exists, and can vary in strength under certain environmental conditions. Although this high-protein allele has been associated with higher seed protein content and a lower yield reduction than have previously characterized high-protein alleles from other sources, there is still much that is not known about the effect of this high-protein allele. Efforts are underway to isolate a high-protein allele at this locus through positional cloning, and at that point it may be possible to understand the gene mechanism and the reason for its sensitivity to different genetic backgrounds and environments.

Both of these projects represent different steps towards the overall process of evaluating and incorporating novel genes from the unadapted soybean germplasm into highly productive elite soybean cultivars. At each stage in this characterization decisions must be made about the
best way to allocate resources. For example, part of this research shows that L20a is highly associated with the Ma reaction; therefore, sources of Ma resistance that are identified in the germplasm in the future could be genotyped with this marker. If any accessions contain the allele associated with Ma resistance at L20a, that source may be set aside for the moment in order to focus on other sources that are Ma resistant but may contain resistance genes on other linkage groups. The finding that L20a is predictive of Ma resistance could also be used within the elite soybean germplasm to select for Ma resistance in early generations of cultivar development, thereby avoiding the process of phenotyping for Ma response, which is a laborious and expensive procedure, until later generations.

One of the final steps in the characterization process is the evaluation of the allele’s effect on yield in different genetic backgrounds. Based on one year of testing, yield was not reduced in lines homozygous for the high-protein allele compared to lines homozygous for the conventional-protein allele, even though seed protein content increased by at least 40 g kg$^{-1}$ in the former. However, the 2-yr averages revealed that the high-protein allele was significantly associated with lower yield in at least two of the three genetic backgrounds. While soybean producers are beginning to receive financial compensation for seeds with higher protein content, most soybean producers still want high-protein and high-yielding soybean cultivars since this would take maximum advantage of a high-protein premium. Since the high-protein allele was shown in this study to be consistently associated with high yield in certain genetic backgrounds but not in others, it may be necessary to verify its effect on each particular genetic background before deploying it on large scale within a soybean cultivar development program. Knowing when and where to allocate resources to best exploit the genetic diversity present in the
germplasm, which is information to which this thesis project contributed, may seem frustratingly slow and at times give contradictory results. However, there is a noble purpose inherent in the process. Upon receiving the Alexander von Humboldt Foundation Award for his significant contribution to American agriculture in the previous 5 years, Steve Tanksley commented, “We've gone back and found wild species that contain genes that may help us boost production. The world is only so big, the population is growing and we need to continue feeding that population.”