Molecular Mapping of Disease-Related Expressed Sequence Tags and Resistance Gene Analogues in Soybean

By

Michael Jason Godwin

Thesis submitted to the faculty of the Virginia Polytechnic Institute and State University

in fulfillment of the requirements for the degree of

Master of Science

In

Crop and Soil Environmental Sciences

Dr. M. A. Saghai Maroof, Chairman

Dr. G. Buss

Dr. S. Tolin

September 18, 2003

Blacksburg, Virginia

Key Words: Disease Resistance, EST, RGA, SSLP Marker Analysis, Integrated Linkage Maps

Molecular Mapping of Disease-Related Expressed Sequence Tags and Resistance Gene Analogues in Soybean

By

Michael Jason Godwin

Dr. M. A. Saghai Maroof, Chairman

Crop and Soil Environmental Sciences

Abstract

Soybean has become one of the most important crops to the United States, resulting in a need to improve its disease resistance. The objectives of this study were to 1) design primers and develop PCR-based markers from disease-related expressed sequence tags (ESTs) and resistance gene analogues (RGAs), 2) assess the utility of such markers by diversity analysis of 12 soybean parental lines, and 3) search for possible association of the markers with known disease resistance genes by constructing a linkage map. The diversity analysis will allow this study to determine how well each marker can distinguish genotypes in soybean. Identifying the location of our markers in the soybean genome with the linkage map, will allow those related to disease resistance to be selected. A total of 202 simple sequence length polymorphism (SSLP) markers were constructed using a set of 1218 disease-related ESTs. Furthermore, 22 markers were constructed using previously identified RGA sequences. Both sets of markers were able to detect polymorphism in the diversity analysis. Also, 48 of the SSLPs, five of the RGAs, and 150 molecular markers were used to construct a soybean linkage map using 114 recombinant inbred lines (RILs). Several markers mapped to chromosomal regions known to contain disease resistance genes. This study has created a framework map, which will be useful for identifying the location of resistance genes, marker-assisted selection for resistance, discovering novel resistance genes, and understanding genome organization of resistance pathways in soybean. An effective approach to develop "candidate gene" markers has been demonstrated.

DEDICATIONS

To my parents, who have unconditionally
sacrificed for my education and well being, and without them
this paper would never have been possible.

And a special thanks to Dr. M. A. Saghai Maroof, for showing me the way.

ACKNOWLEDGEMENTS

I would like to express my gratitude to my advisor, Dr. M. A. Saghai Maroof, and my committee members, Drs. G. Buss and S. Tolin. Dr. Maroof has guided me through my Graduate experience in an exceptional manner. Sternly in a direction to prepare me for a career, but always graciously when I needed someone to pick me up. His expertise provided gains in my learning and final thesis that are irreplaceable. I would also like to thank Drs. G. Buss and S. Tolin for providing advice when it was needed and reviewing my thesis.

Drs. Ruslan Biyashev and Jafar Mammadov have worked extensively with me throughout my research. Without their constant supervision and insights my research would never have been completed. Dr. Biyashev made my understanding of every theoretical concept and laboratory procedure complete, and always gave me his attention unselfishly. Dr. Mammadov took me under his wing and sheltered me through my troubles and hardships. His advice and knowledge were essential to my research.

Once again I would like to thank all my family and friends, especially my parents, who have supported me during my time at Virginia Tech.

TABLE OF CONTENTS

LIST OF TABLES

CHAPTER II.

CHAPTERIII.

LIST OF FIGURES

CHAPTER II.

CHAPTERIII

Chapter I. **Literature Review**

Soybeans, Society, and Genomics

A legume was described in 2838 B.C. by Chinese Emperor Sheng-Nung becoming the first documentation of the existence of the soybean (Morse, 1950). After its discovery, nearly 2000 years passed before its value as a food product was recognized. However, like a fine wine, the importance and taste for this commodity has greatly increased with age. It was first introduced to USA as a forage crop and slowly began to emerge as an oilseed crop. Now, nearly 3 billion bushels are grown annually in the U.S. alone. It has become the world's primary source of seed oil and seed protein (Verma and Shoemaker, 1996). Both people and livestock rely on soybeans as a food source. In addition, its utilization is being expanded to the industrial and pharmaceutical sectors (Smith and Huyser, 1987). Certain derivatives of soybean seed can be used in manufacturing products and have medicinal power. Whenever a plant becomes a focal point of society, its well being and improvement also become important.

The presence of isoflavones in soybean has sparked the interest of clinicians to conduct research on its use in preventing and treating chronic diseases. Several sources have been cited that illustrate isoflavones ability to inhibit growth of cancer cells, lower cholesterol, and prevent bone disorders (Liu, 1997). These characteristics are well suited for developing a treatment for two of the leading causes of death, heart disease and cancer. Many studies have correlated soy intake and the presence of isoflavones with reduced risk to breast cancer, prostate cancer, and kidney disease. There are even speculations that existing tumors can be treated with soy products. One study by Takeda et al. (1994) showed that an isoflavone (genistein) enhanced the effects of chemotherapeutic agents against leukemia cells. These are just a few examples of the uses of soybean in medicine that are being explored. So far, the results and benefits of this approach seem to be promising.

The Chinese have already demonstrated that soybean can be used for purposes other than food. In their culture, protein cakes from crushed seeds have been used as fertilizers for centuries. Further technological research of the protein has indicated a wider range of possible industrial applications, mainly the production of plastics and adhesives. According to Myers (1993), use as an adhesive for wood and paper coatings is the most common. Soy protein can also be found in the ingredients of some plastics and textile fibers, replacing environmental impurities such as fiberglass-reinforced composites. Although soy protein and oil are presently in limited use as industrial resources, they are starting to make their way into a wider range of products each year. Lack of processing technology is probably its main hindrance as an industrial resource. However, due to economical and environmental advantages a few areas that are focusing on increased soy utilization are lubricants, diesel fuels, and pesticides (Johnson, 1984; Mounts et al., 1987; and Liu, 1997). Soybean is a natural, renewable resource, giving it an edge over the many environmental impurities used in these areas.

It should be no surprise that such a valuable commodity has received extensive research towards its productivity and improvement. The obvious goal is to produce as much product with as little cost as possible. This way of thinking has placed an emphasis on soybeans from an agricultural aspect for the past century. Thus, agricultural traits such as crop yield and disease resistance have been given priority.

Understanding and combating the many pathogens that affect soybeans can be crucial to preventing devastating yield losses. Soybeans have one of the most diverse ranges of diseases compared to other crops. More than a hundred different pathogens are known to affect soybeans. Of these documented pathogens, 35 pose serious economic threats (Sinclair and Shurtleff, 1999). All parts of the plant are susceptible to some type of pathogen, which may compound the problem. It is very rare to find a soybean field that is not infected with some pathogen, and in most instances the plants are infected with multiple diseases. A recent study (Wrather et al., 1998) estimated that economic losses in the top ten soybean producing countries due to disease were $6.29 billion.

Traditionally, the main focus of researchers has been to concentrate on improving one trait of a soybean line. It may be resistance to a particular disease or environmental condition, or simply a gene that increases yield in general. Over the last ten years this process has been facilitated with the help of genomics, in particular, genetic linkage maps and molecular marker technology. This new technology increases the efficiency of breeding programs and allows rapid discovery of many agriculturally and economically important genes. It has even gone so far as to allow the direct manipulation of specific genes through cloning and transformation techniques. The rise of this trend can be evidenced by the development of a herbicide resistant soybean line, commonly referred to as "Round-Up Ready Beans". This biotech product has proven to be beneficial and has become a preferred choice by many farmers (Monsanto 2003).

The science has begun to revolutionize the way researchers approach crop improvement. Instead of focusing on one trait to enhance yield alone, a global perspective can be taken that encompasses many traits to strive for an overall superior soybean. The quality of the seed produced becomes just as important as the quantity. The interests of a particular industry can be tailored into a high yielding plant.

Some researchers have gone so far as to target a variety of nutritional characteristics of soy food products. It becomes very difficult to apply traditional breeding strategies when many genes are targeted. Oil and protein content are the major economic traits that are receiving attention. Their concentration can be matched to fit the intended use of the soybean's production. For example, there may be an economic demand for soy foods with high protein and oil. Thus, soybean seed that possesses these two characteristics should be grown, but to reduce costs in processing the seed should also be low in fatty acids. There would also be a social demand for low triacylglycerols and carbohydrates for health reasons. Genomics allows all of these factors to be considered simultaneously in the development of a soybean variety that fulfills the desired needs (Liu, 1997). Biotech applications overcome the traditional limitations that would be encountered. Without molecular markers, addressing any more than one of the mentioned factors would be cumbersome, and most likely unsuccessful.

Genetic Linkage Maps

A genetic linkage map is a physical representation of a genome that shows the positions of specific DNA markers relative to each other. The construction of linkage maps is based solely on the ability to identify genetic markers. Any differentially inherited genetic trait that segregates among progeny is a possible marker. In order to identify a marker, it must be polymorphic. Any variations in the parental DNA of a

particular gene will lead to polymorphism. In general, variations occur within the intron of a gene and have no effect on its function. The most common types of markers used are restriction fragment length polymorphisms (RFLPs) and microsatellites. RFLPs produce different size fragments when cleaved by restriction enzymes because of the variation in the DNA. Microsatellites are tandem repeats in the DNA with variable lengths, depending on the number of repeats in the individual segment.

If a marker can be identified, then its pattern of inheritance can be traced. Mendelian genetics can be used to explain these patterns of inheritance. This is the key principle that all computational programs rely on to construct a genetic linkage map. Linked markers are inherited together; thus, closely linked markers will experience fewer recombination events. A marker's recombination frequency, or percentage of events that separate linked markers, is a function of this concept. Recombination frequencies between two markers decrease as the markers become closer on the chromosome. The basic idea is that the more often two markers are inherited together the closer they are to each other in the genome.

Many computer programs for the analysis of mapping data can be found on the Internet. Most of these programs are reliable and produce similar results. The philosophy behind each program is the same; differences may lie in the weight each variable of the computation is given. Furthermore, certain programs are designed to accept data of a specific type. The goal is to enhance the program's ability to analyze a particular data set. Four widely accepted programs for constructing genetic linkage maps are Mapmaker, Map Manager, MultiMap, and LINKAGE. The main differences between these programs are the type of data they accept, and how far the analysis can be taken.

To create a genetic linkage map, these programs must be able to compute the recombination frequencies of markers in a pedigree data set. This begins with a pairwise comparison of each marker in the data set. The program uses the recombination frequencies computed for each pairwise analysis to assign the markers to linkage groups. Low recombination frequencies mean that two markers are close to each other. Multiple comparisons are then performed on each group of markers. Multipoint analysis also uses recombination frequencies to order the markers in each group. This approach relies on statistics to predict which specific ordering of markers is most likely. The statistical methods chosen to perform all of these computations may differ among mapmaking programs.

There are not many differences between the programs used to construct genetic linkage maps. The final map should be extremely similar regardless of which program is used. Choosing a program is based on personal preference and the goal of the research project. Any interests in the data other than a linkage map should be considered. The key is to find a program that has specific operations designed to explore these interests.

Linkage maps have become so appealing due to their resourcefulness. Molecular markers are very informative alone, and when they are compiled into one map the opportunities for their application increase. Molecular techniques and marker technology have been directed towards studying genes with agriculturally or economically important traits. Genetic linkage maps have been the cornerstone for this goal over the past decade. They provide information for analysis of quantitative trait loci (QTLs), understanding genetic variation in germplasm collections, gene tagging, gene cloning, and marker-

assisted selection (MAS) (Motto and Marsan, 2002). The information is there, how it will be applied will determine what questions can be answered.

Many crops, such as maize, have already experienced the benefits of linkage maps. The first simple sequence repeat-based published map of maize was by Senior et al. (1997). Other researchers soon followed suit, constructing their own linkage maps in crops they were studying. The intent was to create separate compositions of marker information that could be pooled and compared for greater understanding in areas such as gene organization, genome evolution, targeted cloning, and interpretation of complex traits (Motto and Marsan, 2002). The results have been more than rewarding for many cropping systems. QTLs have been identified for grain yield (Ajmone-Marsan et al., 2001), forage yield (Lubberstedt et al., 1997), herbicide tolerance (Sari-Gorla et al., 1997), insect resistance (Bohn et al., 1996), and pathogen resistance (Holland et al., 1998). The list of agronomically important traits discovered goes on, these are just a few examples.

Genetic linkage maps that produce tightly linked markers can be very beneficial to marker assisted selection (MAS) breeding programs. In this technique, researchers employ molecular markers to improve the efficiency of selecting for genes of interest. This application has been established well in rice, tomato, wheat, and barley (Motto and Marsan, 2002). Successful programs implement the strategy to efficiently pyramid resistance genes into high yielding varieties. A few landmark achievements that relied upon molecular marker information for gene pyramiding have been reported including, blast (Hittalmani et al., 2000) and gall midge (Katiyar et al., 2001) resistance in rice. MAS work is also currently being introduced into soybeans for soybean mosaic virus resistance (Buss, Maroof, and Tolin, Unpublished).

If a map can be saturated with enough of these tightly linked markers for traits of interest, then gene cloning becomes possible. Martin et al. (1993) illustrated this idea by cloning a protein kinase gene (*Pto*) that confers resistance to bacterial speck in tomato. Tomatoes have been studied for well over fifty years, resulting in numerous dense, informative genetic linkage maps. The researchers took advantage of these maps to identify markers that span the region of the genome containing the *Pto* gene. The markers were then used to scan a yeast artificial clone (YAC) library of tomato to identify a clone containing the resistance locus. A cDNA clone of the *Pto* gene could then be isolated from the YAC clone.

The process of map-based cloning has been accelerated in such crops as tomato due to a small genome size and low levels of repeated sequences. However, this does not exclude the more complex genomes from efficiently utilizing this technology. Gene cloning has already been demonstrated in soybean (*Rsv1*, gene that confers resistance to soybean mosaic virus), a genome that has forty to sixty percent repetitive sequences (Saghai Maroof, unpublished). The foundation for cloning of *Rsv1* was laid from research done by Gore et al. (2002). The researchers identified a disease resistance gene cluster on soybean linkage group F. *Rsv1* and *Rpv1* were associated with this region. Three tightly linked markers were identified for these loci from a previous study (Hayes et al., 2000). Fine mapping was then used to create a high-resolution genetic map of the chromosomal region around these loci. The intent was to provide tightly linked disease resistance related markers that could be used for cloning purposes. The efforts of the

study were beneficial and the ultimate goal of gene cloning was achieved (Gore et al., 2002).

Comparative mapping is yet another tool of genetic linkage maps that has proven to be very effective. The idea is to develop genome maps of similar species using a common set of molecular markers. This allows the organization of chromosomes among related species to be studied. Information about the evolution and genetic structure of a species can be obtained. Synteny, identical gene organization, can be used to expedite gene cloning and genome mapping. Rice-barley synteny was useful for the rapid cloning of a stem rust resistance gene in barley using markers from a saturated rice map (Brueggeman et al., 2002; Han et al, 1999; Kilian et al., 1997). It has also been demonstrated in a study by Ahuja et al. (1994) that a single RFLP probe was capable of hybridizing to a variety of conifer species. The principal hope of comparative mapping is to facilitate the genetic research of all related species. Saghai Maroof et al. (1996) illustrated a high degree of synteny between rice and barley using RFLPs. Their results allowed them to uncover the possibility of developing one representative map based on genome homoeologies for the *Gramineae* plant family, which includes many important cereal crops such as rice, barley, wheat, and maize.

In a study by Devey et al. (1999), a common set of RFLP markers was used to create maps in loblolly and radiata pine. The maps were highly collinear and provided a basis for the identification of homologous regions that may represent QTLs for economically important characteristics. The result of such high similarity among the markers allows gene mapping in both species to be accelerated. A marker linked to a gene from one map can be used to predict the location of the same gene on the other map. The researchers believe that the collinearity is high enough that a "generic" pine map could be produced. The information from one map could be conveyed to all pine species. Generic maps provide a vast supply of heterologous markers for specific locations in the genome that can be used to enhance linkage map construction and positional cloning in species less studied (Saghai Maroof et al., 1996).

Molecular Markers

Genetic variability that can be detected at the molecular level has been a monumental break through for crop improvement research. Plant breeding programs are employing molecular markers. Breeders are no longer hindered by relying solely on observable traits as selection markers. Phenotypes correlated with many traits of interest are limited and often affected by many environmental factors that make their identification difficult and unreliable. The ability to analyze genetics at the DNA level provides accuracy and the possibility to exploit all genomic diversities. Molecular marker technology has greatly improved the ability to identify clones, breeding lines, hybrids, and cultivars, to track transformed DNA, to assess genetic diversity in a germplasm collection, and to perform phylogenetic analysis (Gupta et al, 1999; Jain et al. 2002).

There are also commercial advantages to using molecular markers. The issues addressed are social concerns mainly dealing with law and policy. Lee and Henry (2001) report that product quality assurance and support for protection of novel plant varieties are achieved with molecular markers. Genotyping allows the exact identification of a

plant. It has been demonstrated that morphologically similar plant varieties have been differentiated using molecular profiling (Yang et al., 1994; Graham et al., 1994; Ko et al., 1996; Scott et al., 2000a). Consumers can be guaranteed about the fidelity and uniformity of the plant variety they have purchased. Plant breeders can also use genotyping to patent unique varieties they have developed (Lee and Henry 2001).

Molecular markers can be categorized into two broad classes, hybridization-based DNA markers and PCR-based DNA markers. The primary purpose of all markers is to create detailed genetic linkage maps that can be employed for gene research. There are advantages and disadvantages to each type of marker that exists. The goal is to recognize which characteristics are best suited for the study at hand.

Hybridization-based Markers

This category of markers refers mainly to restriction fragment length polymorphisms (RFLPs), and was the first type of molecular marker to be developed. The polymorphisms detected are based on accumulated mutations in the genome, such as substitutions or deletions. When the genome is digested with restriction enzymes the mutations will produce different size fragments. Gel assay, Southern blotting, and radioactive or non-radioactive labeled probes can be utilized to detect RFLPs (Yu et al., 1994). This type of marker is co-dominant, highly reproducible, and can be used to characterize DNA. Saturated genetic linkage maps have been constructed in tomato, rice, wheat, maize, lettuce, and barley using RFLPs (Winter et al, 2002).

The work required for this method is time consuming and labor-intensive, rendering it less useful for rapid evaluation in large-scale studies. Large amounts of DNA are required and the cost of this procedure can become expensive when compared to PCR-based markers. Species-specific probes are also required to detect the markers, which add to the constraints of its application.

PCR-based Markers

PCR-based markers include a wide range of techniques that differ mainly in sequences and position of primers utilized. A few common types implemented include random amplified polymorphic DNAs (RAPDs), simple sequence repeats (SSRs), and amplified fragment length polymorphisms (AFLPs). DNA segments can be amplified using primers that anneal to flanking sequences. The segments can be gel assayed and detected with either ethidium bromide staining or radioactive labeling (Yu et al., 1994). Primers can be designed for specific sequences, which allow precise DNA regions of interest to be targeted. These types of markers offer a potential for increased efficiency, while reducing expense. PCR markers have become a preferred choice in marker aided selection programs, quality assurance cases, and tracking of transgenes (Brar, 2002; Lee and Henry, 2001).

The recent explosion of available DNA sequence information resulting from large scale sequencing projects has greatly benefited PCR-based markers. Sequences are compared and scanned for differences to design primers that flank these polymorphic regions. The technique allows for detailed analysis of the sequence and specific nucleotide differences can be detected. The development of new marker types such as

single nucleotide polymorphisms (SNPs) and simple sequence length polymorphisms (SSLPs) have stemmed from improvements of this application. SNPs are biallelic markers that are designed based on a single nucleotide difference in DNA sequence. Their use has been instrumental for rapid and highly automated genotyping in humans and is expected to bring the same accomplishments to plant genomics (Winter et al., 2002; Zhu et al., 2003). SSLPs are designed from differences identified across short segments of the DNA. These markers are polyallelic and more abundant, making them more desirable than SNPs (Gupta et al., 1999). This review will focus on SSLPs, some of the more common PCR-based markers mentioned above will not be covered in detail. SSRs will be included in the discussion because of their utility in SSLPs.

Microsatellites

The term microsatellite, or simple sequence repeat (SSR), refers to tandem repeated sequences of variable length. Repetitive sequences are easily mutated during replication, due to polymerase slippage, and result in highly polymorphic regions (Guyomarc'h et al., 2002). Generally, a repeat is gained or lost which produces variable lengths at that locus. SSRs appear to be randomly dispersed through the genome and are generally flanked by conserved regions (Winter et al., 2002). The characteristics are well suited for amplification as PCR products, leading to their extensive development as molecular markers.

Expressed Sequence Tags (ESTs)

ESTs offer a more informative type of sequence information that is ideal for functional genomics. Their sequences are developed from cDNA clones that correspond to mRNA. There is no doubt that the sequence is a functional gene that was turned on under the conditions in which the mRNA was isolated. Many genes can be analyzed under many conditions, allowing the interaction of genes during biological processes to be monitored. The sequences are derived from conserved domains of the genome and should possess excellent transferability of linkage maps across species. For example, a barley set of ESTs has been used to identify salt-induced genes in foxtail millet (Sreenivasulu et al., 2001). The combined characteristics of ESTs are ideal for molecular techniques such as microarray analysis, genetic linkage map construction, and map-based cloning, all practices devoted to gene discovery and characterization. Human and mouse genomes have already demonstrated the usefulness of ESTs. However, their application in plant species has been limited. When compared to mammalian genomes, there is considerably fewer plant ESTs in available databases (Wang et al., 2001).

Efforts are currently being made in plant genomics to establish a publicly available database, GenBank, which contains a large number of ESTs. Shoemaker et al. (2002), compiled more than 120,000 soybean ESTs for analysis and organization. Bioinformatics tools such as sequence analysis and annotation, contig assembly, and cluster analysis were applied to identify genes with similar expression patterns, and possibly their function. Their results were used to create a well-organized, public database that grouped related ESTs. Utilization of such databases allows the structure, function, and evolution of a model crop genome to become attainable (Shoemaker et al.,

2002). Currently, soybean leads all crop plants with over 250,000 ESTs in GenBank (Shoemaker et al., 2002).

SSRs as a Source for Molecular Markers

Microsatellite markers are generated as PCR products that span regions containing tandem repeats. Primers can be designed flanking microsatellite repeats. Theoretically, the technique is based on creating a PCR product that has a length directly related to the length of the microsatellite it flanks. Thus, microsatellites that are polymorphic will produce PCR products with different sizes. Genomic DNA libraries have commonly been used as sources. Rather than using random DNA sequences, markers can be generated from within functional genes. Therefore, many researchers are turning to cDNA libraries to obtain sequences for marker design. The trend in current studies is the application of ESTs for marker design, and this will be the only approach discussed in this review.

The nature of polymorphism in SSRs gives them several distinct advantages over hybridization-based methods, such as RFLPs. Due to the detection of multiple loci, the identification of a single locus using RFLPs becomes extremely difficult in crops such as soybean. Unique bands may segregate depending on the mapping population being used. In order to identify a RFLP locus, more considerations must be taken into account. The probe, restriction enzyme being used, and molecular weight of the segregating band must all be defined. These complexities make comparing linkage maps difficult. Adding to the problem is the low level of polymorphism detected by RFLPs. In soybean, RFLPs rarely detect more than two alleles at a locus (Cregan et al., 1999). The chance of identifying multiple variant alleles is greatly reduced. Thus, a marker's use will be limited to one, or just a few, genetic linkage maps. SSRs on the other hand, detect polymorphism at a single locus. Fewer amplification products will be generated and they will be more specific. Band analysis will be simplified with fewer PCR products. This reduces the complexity of scoring the bands leading to fewer mistakes and easier reproducibility of the marker.

SSR markers also exhibit codominant segregation which allows multiallele identification. Dominant markers detect polymorphism based on the presence or absence of a band. Polymorphism is detected in SSRs by characterizing the different sizes of the bands produced. As long as an allele produces a gene fragment with a different size, SSRs can detect it. The ability to visualize both parental bands also allows heterozygosity to be monitored which is not possible when scoring based on presence or absence alone.

One of the most appealing features of SSRs is their high rate of polymorphism. This stems from their ability to detect multiple alleles at one locus. It was demonstrated early on by the work of Akkaya et al. (1992), that SSRs possessed a highly polymorphic nature in soybeans**.** Saghai Maroof et al. (1994) have also illustrated the highly polymorphic rate of microsatellites in barley. They were able to detect 28 and 37 variant alleles at two different loci, which is the largest number of alleles ever reported in plants for single Mendelian loci. Detection at this scale improves the chances of parental lines being polymorphic. Multi-parental polymorphism allows researchers to accomplish more with one marker. Genetic linkage maps of different parental crosses can be associated to

study chromosome homology and give a better overall understanding of gene order. Even more specifically, the fate of genomic segments can be tracked (Cregan et al., 1999).

The occurrence of SSRs throughout the genome seems to be ideal. Not only are they abundant, but they seem to be randomly distributed. It is reported by Arahana et al. (2001), that an SSR occurs every 10 kilobases in the soybean genome. Human and mammalian research has also illustrated the randomness of SSRs in genomes. These findings about the distribution of SSRs in DNA were previously supported by Akkaya et al. (1995), who found minimal clustering of SSRs in the soybean genome.

The initial cloning and sequencing of DNA that is required to produce SSR markers can be costly. However, once a primer becomes available for a specific microsatellite, the marker becomes very cost-effective (Winter et al., 2002). Data can be collected rapidly for these markers, and the results are highly reproducible. Most effort has been directed toward the development of saturated genetic linkage maps. However, they are suitable for all types of molecular marker work.

Cregan et al. (1999) have successfully developed an integrated genetic linkage map of the soybean genome using three different mapping populations. SSRs were chosen over RFLPs for the following two reasons: 1) RFLPs rarely detect more than two alleles at one locus and 2) one probe will generally detect multiple loci. Both reasons make it difficult to detect high levels of polymorphism in parental lines, limiting possible mapping populations for the markers. Thus, fewer linkage maps will be able to use the RFLPs. The experiment resulted in the identification and alignment of 20 homologous linkage groups, representing the 20 chromosomes of soybeans, between the three maps.

The efficiency of SSRs for MAS of QTLs has been demonstrated by Li et al. (2001), in conditioning soybean resistance to southern root-knot nematode. The abundance and highly polymorphic nature of SSRs that have become publicly released allowed the researchers to quickly identify several markers tightly flanking the QTL. The comparison of published maps for markers near the region of interest accelerated the process. The linked markers were then used to rapidly breed the resistance genes into a productive soybean line. Similar procedures for MAS have also been successful with soybean mosaic virus (SMV) and soybean cyst nematode (SCN) (Dr. Buss and Dr. Tolin, unpublished). The use of SSRs for tagging disease resistance genes has been very successful in soybeans. Examples include resistance to soybean mosaic virus (SMV) (Gore et al., 2002; Jeong et al., 2002) and sudden death syndrome (SDS) (Iqbal et al., 2001).

ESTs as a Source for Molecular Markers

Predicting the location of a gene and map comparison are two of the most desired qualities of markers. The closer a marker is to a gene, the better it serves to tag the gene. The ability to transfer more markers to a greater number of maps will greatly improve information that can be obtained from map comparison. SSR markers derived from ESTs offer both. The sequences that are being mapped are part of a functional gene. It is possible that one of the markers is part of a gene of interest. It has also been reported that sequences flanking SSRs within ESTs are highly conserved (Cato et al., 2001; Eujayl et

al, 2002). This could lead to a greater chance of marker transferability. A conserved SSR will have more opportunities to be used as a marker in a greater number of genomes.

Marker design from ESTs is identical to PCR marker development as described above. The sequences are scanned for polymorphic regions and primers are designed to flank these regions. Any one of the PCR-based methods can be applied. According to the literature, SSRs seem to be a good approach that results in high polymorphism and transferability (Scott et al., 2000b; Eujayl et al., 2002). The excellent distribution of SSRs found in genomic sequences seems to carry over to ESTs. It was determined in a 24,344 wheat and barley EST collection, by Holton et al. (2002), that SSRs accounted for nearly 6% of the sequences as a whole. The availability of repeats in ESTs increases the chances of polymorphisms, possibly leading to the development of more markers. SSLP application can be combined with SSR development when scanning the sequences for polymorphism. This allows marker construction to also include random differences within the sequences, such as gaps.

The GenBank database, published by NCBI, can be easily accessed from their homepage (//www.ncbi.nlm.nih.gov). Homologous ESTs can be pooled for sequence analysis with a few clicks of a mouse button, allowing the time and expense of developing PCR-based markers to be greatly reduced. High-throughput methods for genetically mapping ESTs can be implemented (Cato et al., 2001). This approach has been successfully used for grapes and rice, and is being applied to other crops such as wheat, barley, and soybean (Winter et al., 2002; this thesis).

In a study by Scott et al. (2000b), analysis was performed on SSRs derived from grape ESTs. Sixteen primer pairs were designed and ten produced an amplified product in an analysis of seven grape cultivars. All ten detected polymorphism and could be used for identity determination of the cultivars. Four amplified several alleles from a related genus demonstrating that ESTs are transferable at a taxonomic level. The findings of this study have also been supported by research conducted in wheat (Eujayl et al., 2002).

Wang et al. (2001) took advantage of DuPont's rice EST database to identify 162 disease-related sequences. A doubled haploid mapping population was used to map the ESTs by RFLP analysis. Map location was determined for 109 of the genes, and clustering was observed at several chromosomal regions. Map comparison from previous studies determined that three of the clusters were associated with QTLs for resistance to various diseases. It is very possible that these ESTs are parts of the actual genes that convey resistance. Wang's study provides excellent support for the candidate-gene approach in gene identification.

SSLP Markers

Simple sequence length polymorphism (SSLP) application encompasses and expands on SSR marker development that has been discussed in this review. Sequence information is scanned for sites that may uncover polymorphism. This scheme includes, but is not limited to, SSR type of polymorphism. It also includes any insertion or deletion that alters the base pair length of the DNA sequence. For instance, when a short segment of DNA is replicating several base pairs could be deleted resulting in a gap between this sequence and a homologous sequence at this location. The polymorphic

region is identified and amplified with primers the same as SSRs that are identified to have a different number of repeats.

The SSLP strategy has already been successfully implemented in rice. Researchers have been able to use the new application for genotype identification and quality assurance of commercial lines (Bligh et al., 1999; Vaughan et al., 2001) and gene and QTL evaluations (Moncada et al., 2001). SSLPs share the same advantages over RFLP analysis as SSRs. According to Bligh et al. (1999), after primers are designed for specific SSLP loci, they are also cheaper, faster, and more tolerant of poor quality DNA than AFLP analysis.

References

Ahuja, M.R., Devey, M.E., Groover, A.T., Jermstad, K.D., and Neale, D.B. 1994. Mapped probes from loblolly pine can be used for restriction fragment length polymorphism mapping in other conifers. Theor. Appl. Genet. **88**: 279-282.

Ajmone-Marsan, P., Gorni, C., Chitto, A., Redaelli, R., van Vijk, R., Stam, P., and Motto, M. 2001. Identification of QTLs for grain yield and grain-related traits of maize (*Zea mays L.)* using AFLP map, different testers, and cofactor analysis. Theor. Appl. Genet. **102**: 230-243.

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

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

Arahana, V.S., Graef, G.L., Specht, J.E., Steadman, J.R., and Eskridge, K.M. 2001. Identification of QTLs for resistance to *Sclerotinia sclerotiorum* in soybean. Crop Sci. **41**: 180-188.

Bligh, H.F.J., Blackhall, N.W., Edwards, K.J., and McClung, A.M. 1999. Using amplified fragment length polymorphisms and simple sequence length polymorphisms to identify cultivars of brown and white milled rice. Crop Sci. **39**: 1715-1721.

Bohn, M., Khairallah, M.M., Gonzalez-de-Leon, D., Hoisington, D., Utz, H.F. Deutsch, J.A., Jewell, D.C., Mihm, J.A., and Melchinger, A.E. 1996. QTL mapping in tropical maize: I. Genomic regions affecting leaf feeding resistance to sugarcane borer and other traits. Crop Sci. **36**: 1352-1361.

Brar, D.S. 2002. Molecular marker assisted breeding. p.55-84. *In* M. S. Jain (ed.) Molecular techniques in crop improvement. Kluwer Academic Publishers, Norwell, MA.

Brueggeman R., Rostoks, N., Kudrna, D., Kilian, A., Han, F., Chen, J., Druka, A., Steffenson, B., and Kleinhofs, A. 2002. The barley stem rust-resistance gene *Rpg1* is a novel disease-resistance gene with homology to receptor kinases. PNAS **99**: 9328-9333.

Cato, S.A., Gardner, R.C., Kent, J., and Richardson, T.E. 2001. A rapid PCR-based method for genetically mapping ESTs. Theor. Appl. Genet. **102**: 296-306.

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

Devey, M.E., Sewell, M.M., Uren, T.L., and Neale, D.B. 1999. Comparative mapping in loblolly and radiata pine using RFLP and microsatellite markers. Theor. Appl. Genet. **99**: 656-662.

Eujayl, I., Sorrells, M.E., Baum, M., Wolters, P., and Powell, W. 2002. Isolation of EST-derived microsatellite markers for genotyping the A and B genomes of wheat. Theor. Appl. Genet. **104**: 399-407.

Gore, M.A., Hayes, A.J., Jeong, S.C., Yue Y.G., Buss, G.R. and Saghia Maroof, M.A. 2002. Mapping tightly linked genes controlling potyvirus infection at the *Rsv1* and *Rpv1* region in soybean. Genome. **45**: 592-599.

Graham, G.C., Henry, R.J., and Redden, R.J. 1994. Identification of navy bean varieties using random amplification of polymorphic DNA. Australian Journal of Experimental Agriculture **34**: 1173-1176.

Gupta, P.K., Varshney, R.K., Sharma, P.C., and Ramesh, B. 1999. Molecular markers and their applications in wheat breeding. Plant Breeding. **118**: 369-390.

Guyomarc'h, H., Sourdille, P., Charmet, G., Edwards, K.J., and Bernard, M. 2002. Characterization of polymorphic microsatellite markers from *Aegilops tauschii* and transferability to the D-genome of bread wheat. Theor. Appl. Genet. **104**: 1164-1172.

Han, F., Kilian, A., Chen, J.P., Kudrna, D., Steffenson, B., Yamamoto, K., Matsumoto, T., Sasaki, T., and Kleinhofs, A. 1999. Sequence analysis of a rice BAC covering the syntenous barley *Rpg1* region. Genome. **42**: 1071-1076.

Hayes, A.J., Yue, Y.G., and Saghai Maroof, M.A. 2000. Expression of two soybean resistance gene candidates shows divergence of paralogous single-copy genes. Theor. Appl. Genet. **101**: 789-795.

Hittalmani, S., Parco, A., Mew, T.V., Zeigler, R.S., and Huang, N. 2000. Fine mapping and DNA marker-assisted pyramiding of the three major genes for blast resistance in rice. Theor. Appl. Genet. **100**: 1121-1128.

Holland, J.B., Uhr, D.V., Jeffers, D., and Goodman, M.M. 1998. Inheritance of resistance to southern corn rust in tropical-by-corn-belt maize populations. Theor. Appl. Genet. **96**: 232-241.

Holton, T.A., Christopher, J.T., McClure, L., Harker, N., and Henry, R.J. 2002. Identification and mapping of polymorphic SSR markers from expressed gene sequences of barley and wheat. Molecular Breeding **9**: 63-71.

Iqbal, M.J., Meksem, K., Njiti, V.N. Kassem, M.Y., and Lightfoot, D.A. 2001. Microsatellite markers identify three additional quantitative trait loci for resistance to soybean sudden-death syndrome (SDS) in essex x forrest RILs. Theor. Appl. Genet. **102**: 187-192.

Jain S.M., Brar, D.S., Ahloowalia, B.S. 2002. Molecular techniques in crop improvement. p. 616. Kluwer Academic Publishers, Boston.

Jeong, S.C., Kristipati, S., Hayes, A.J., Maughan, P.J. Noffsinger, S.L., Gunduz, I., Buss, G.R., and Saghai Maroof, M.A. 2002. Genetic and sequence analysis of markers tightly linked to the soybean mosaic virus resistance gene, *Rsv3*. Crop Sci. **42**: 265-270.

Johnson, D. 1984. Nonfood Uses of Soy Protein Products. p.175-181. In R. Shibles (ed.) World Soybean Research Conference III. Westview Press, Inc., Boulder, Colorado.

Katiyar, S.K., Tan, Y., Huang, B., Chandel, G., Xu, Y., Zhang, Y., Xie, Z., and Bennett, J. 2001. Molecular mapping of gene *Gm6(t)* which confers resistance against four biotypes of Asian rice gall midge in China. Theor. Appl. Genet. **103**:953-961.

Kilian, A., Chen, J., Han, F., Steffenson, B., Kleinhofs, A. 1997. Towards map-based cloning of the barley stem rust resistance genes *Rpg1* and *rpg4* using rice as an intergenomic cloning vehicle. Plant Mol. Biol. **35**: 187-195.

Ko, H.L., Henry, R.J., Beal, P.R., Moisander, J.A., Fisher, K.A. 1996. Distinction of *Ozothamnus diosmifolius* (Vent.) DC genotypes using RAPD. HortScience **31**: 858-861.

Lee, L.S., and Henry, R.J. 2001. Commercial applications of plant genotyping: In: Plant genotyping: the DNA fingerprinting of plants. Ed. Henry, R.J. CAB International.

Li, Z., Jakkula, L., Hussey, R., Tamulonis, J., and Boerma, H. 2001. SSR mapping and confirmation of the QTL from PI96354 conditioning soybean resistance to southern root-knot nematode. Theor. Appl. Genet. **103**: 1167-1173.

Liu, K. 1997. Soybeans: Chemistry, Technology, and Utilization. Chapman and Hall. New York, New York.

Lubberstedt, T., Melchinger, A.E., Schon, C., Utz, H.F., and Klein, D. 1997. QTL mapping in testcrosses of European Flint lines of maize. I. Comparison of different testers for forage yield traits. Crop Sci. **37**: 921-931.

Martin, G.B., Brommonschenkel, S.H., Chunwongse, J., Frary, A., Ganal, M.W., Spivey, R., Wu, T., Earle E.D., and Tanksley S.D. 1993. Map-based cloning of a protein kinase gene conferring disease resistance in tomato. Science **262**: 1432-1436.

Moncada, P., Martinez, C.P., Borrero, J., Chatel, M., Gauch, H., Guimaraes, E., Tohme, J., McCouch, S.R. 2001. Quantitative trait loci for yield and yield components in an Oryza sativa x Oryza BC_2F_2 population evaluated in an upland environment. Theor. Appl. Genet. **102**: 41-52.

Monsanto Company. 2003. "About Monsanto." Australia: Monsanto Company, retrieved September 30, 2003 http://www.monsanto.com.au.

Morse, W.J. 1950. History of soybean production. p.3-57. *In* K. S. Markley (ed.) Soybeans and soybean products. vol 1. Interscience Publishers, Inc., New York. 2 vols.

Motto, M., and Marsan, P.A. 2002. Construction and use of genetic maps in cereals. p.347-370. In M. S. Jain (ed.) Molecular Techniques in Crop Improvement. Kluwer Academic Publishers. Norwell, MA.

Mounts, T.L., Wolf, W.J., and Martinez, W.H. 1987. Processing and Utilization. p.819-866. In J. R. Wilcox (ed.) Soybeans: Improvement, Production, and Uses. Second edition. American Society of Agronomy, Madison, Wisconsin.

Myers, D. J. 1993. Industrial application for soy protein and potential for increased utilization. Cereal Foods World. **38**:355.

NCBI. 2003. "National Center for Biotechnology Information." Maryland: National Center for Biotechnology Information, retrieved September 30, 2003 http://www.ncbi.nlm.nih.gov.

Saghai Maroof, M.A., Biyashev, R.M., Yang, G.P., Zhang, Q., and Allard, R.W. 1994. Extraordinarily polymorphic microsatellite DNA in barley: species diversity, chromosomal locations, and population dynamics. PNAS **91**: 5466-5470.

Saghai Maroof, M.A., Yang, G.P., Biyashev, R.M., Maughan, P.J., and Zhang, Q. 1996. Analysis of the barley and rice genomes by comparative RFLP linkage mapping. Theor. Appl. Genet. **92**: 541-551.

Sari-Gorla, M., Krajewski, P., Binelli, G., Frova, C., Taramino, G., and Villa, M. 1997. Genetic dissection of herbicide tolerance in maize by molecular markers. Molecular Breed. **3**: 481-493.

Scott, K.D., Ablett, E.M., Lee, L.S. and Henry, R.J. 2000a. AFLP markers distinguish an early mutant of Flame Seedless grape. Euphytica. **113**: 245-249.

Scott, K.D., Eggler, P., Seaton, G., Rossetto, M., Ablett, E.M., Lee, L.S., and Henry, R.J. 2000b. Analysis of SSRs derived from grape ESTs. Theor. Appl. Genet. **100**: 723-726.

Senior, M.L., Chin, E.C.L., Lee, M., Smith, J.S.C., and Stuber, C.W. 1997. Simple sequence repeat markers developed from maize sequences found in the GENBANK database: map construction. Crop Sci. **36**: 1767-1783.

Shoemaker, R., Keim, P., Vodkin, L., Retzel, E., Clifton S., Waterston, R., Smoller, D., Coryell, V., Khanna, A., Erpelding, J., Gai, X., Brendel, V., Raph-Schmidt, C., Shoop, E., Vielweber, C., Schmatz, M., Pape, D., Bowers, Y., Theising, B., Martin, J., Dante, M. Wylie, T., and Granger, C.,. 2002. A compilation of soybean ESTs: generation and analysis. Genome **45**: 329-338.

Sinclair, J.B., and Shurtleff, M.C. 1999. Compendium of Soybean Diseases. (4thed.). The American Phytopathological Society, Inc. St. Paul, Minnesota.

Smith, K. J., and Huyser, W. 1987. World Distribution and Significance of Soybean. p.1-22. In J. R. Wilcox (ed.) Soybeans: Improvement, Production, and Uses. Second Edition. American Society of Agronomy, Madison, Wisconsin.

Sreenivasulu, N., Wobus, U., Mirand, M., and Weschke, W. 2001. A set of salt-induced genes identified in tolerant foxtail millet seedlings by using barley cDNA macroarray. 4th International triticeae symposium, September 10-12. Cordoba, Spain. pp.137.

Takeda, Y., Nishio, K., Nhtani, H., and Saijo, N. 1994. Reversal of multidrug resistance by tryosine-kinase inhibitors in a non-P-glycoprotein-mediated multidrug-resistant cell line. Int. J. Cancer 57:229.

Vaughan, L., Ottis, B., Prazak-Harvey, A., Sneller, C., Chandler, J., and Park, W. 2001. Is all red rice found in commercial rice really Oryza sativa? Weed Science **49**: 468-476.

Verma, D.P.S., Shoemaker, R.C. 1996. Soybean: Genetics, Molecular Biology and Biotechnology. p.270. CAB International, UK.

Wang, Z., Taramino, G., Yang, D., Liu, G., Tingey, S.V., Miao, G.H., and Wang, G.L. 2001. Rice ESTs with disease-resistance gene or defense-response gene-like sequences mapped to regions containing major resistance genes or QTLs. Mol. Genet. Genomics **265**: 302-310.

Winter, P., Huttel, B., Weising, K., and Kahl, G. 2002. Microsatellites and molecular breeding: exploitation of microsatellite variability for the analysis of a

monotonous genome. p.85-138. In M. S. Jain (ed.) Molecular Techniques in Crop Improvement. Kluwer Academic Publishers. Norwell, MA.

Wrather, J.A., Anderson, T.R., Arsyad, D.M., Tan, Y., Ploper, L.D., Porta-Puglia, A., Ram, H.H., and Yorinori, J.T. 1998. Soybean disease loss estimates for the top ten soybean producing countries in 1998. The Canadian Journal of Plant Pathology. **23**: 115-121.

Yang, G.P., Saghai Maroof, M.A., Xu C.G., Zhang, Q., and Biyashev, R.M. 1994. Comparative analysis of microsatellite DNA polymorphism in landraces and cultivars of rice. Mol. Gen. Genet. **245**: 187-194.

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

Zhu, Y., Song, Q., Hyten, D., Van Tassell, C., Matukumalli, L., Grimm, D., Hyatt, S., Fickus, E., Young, N., and Cregan, P. 2003. Single-nucleotide polymorphism in soybean. Genetics **163**: 1123-1134.

Chapter II. **Diversity Analysis of Soybeans using Disease-Related ESTs and RGAs**

Abstract

Soybean has emerged as one of the most important crops in the United States. Molecular marker technology has become a primary instrument in crop research and improvement. Thus, enhancing the types of markers that can be applied in soybean has become a top priority. The objectives of this study were to (1) assess the extent of SSLP variation in soybean ESTs, (2) determine the extent of variation in soybean RGAs, and (3) examine the suitability of ESTs and RGAs for PCR marker analysis. In 1218 disease-related ESTs, 202 SSLPs were identified and used to design PCR markers along with 22 previously identified RGAs. A diversity analysis was employed to assess the markers' ability to distinguish variant alleles in twelve differential soybean lines that represented wild and cultivated lines. Amplification products were observed in 187 of the SSLPs and 21 of the RGAs. In the analysis, 66 of the 187 SSLPs and 6 of the 21 RGAs detected polymorphism among the 12 soybean lines. These polymorphic markers identified 213 SSLP and 18 RGA variants. Genetic diversity values were calculated to assess the ability of each marker to differentiate genotypes. The polymorphic nature exhibited by each marker suggests an increased chance of being useful across various soybean germplasm. Several SSLPs were identified with sufficiently high diversity values that they would make excellent candidates for integration into many different mapping projects. The overall results with the two types of markers were compared and found to be very similar. A high level of polymorphism in soybean was detected for the SSLPs, identified in ESTs, and RGAs. Both types of markers were successful at differentiating variant alleles between the twelve soybean lines. This study demonstrates the utility of using ESTs and RGAs as a source for PCR markers.

Introduction

Soybean has risen to an elite status as a crop because of its broad range of utilization and production. It is the leading oilseed crop in the world (Riley, 1999) and is also expected to become the number one food source (Liu, 1997). In addition, industrial manufacturers are beginning to exploit soy products for adhesives, textile fibers, fuels, and medicinal purposes. Soybean has become a multibillion-dollar crop for the United States, placing enormous importance on the improvement of its agronomic and economic traits. The complexity of gene interactions that govern desirable soybean characteristics has led to the emergence of molecular markers as a premier tool for developing cultivars for use in this cropping system. Soybean genomics has experienced advancement with the construction of a saturated genetic linkage map (Cregan et al., 1999), the cloning of disease resistance genes (Gore et al., 2002; Hayes et al., in press), and its capacity for genetic transformation (Clemente et al., 2000). Marker technology was vital to each of these accomplishments, which makes its enhancement crucial for future progress.

Using ESTs as a source for marker construction increases their potential for more insightful information. It is possible for markers to be designed from sequences contained within genes of interest. As a result, gene discovery and cloning can greatly be facilitated (Sterky et al., 1998; Wang et al., 2001). PCR-based marker application utilizing ESTs has successfully been implemented in wheat, barley, grapes, and rice (Holton et al., 2002; Scott et al., 2000; Wang et al., 2001). Public databases are increasing in the number of ESTs available daily. It was reported in 2002 that over 250,000 soybean ESTs were publicly accessible, the most of all crop plants (Shoemaker et al., 2002). With EST information being so easily attainable, the cost and time of marker development is greatly reduced. We will refer to such EST-derived markers as simple sequence length polymorphism or SSLPs, hereafter. Using SSLPs to construct markers is very similar to SSR technology. The majority of the markers detect differences in microsatellites at a particular locus. However, primer design is not limited to spanning repeat regions. SSLPs include all possible short sequence differences that may arise at a locus creating multiple alleles, such as insertion/deletion mutations that result in sequence gaps between alleles. The benefits of SSR application are being expanded to include more types of short sequence polymorphism, which should result in a larger number of markers produced.

This improved strategy has already been successfully applied to some crop species such as, rice (Moncada et al., 2001; Cho et al., 1998) and soybean (Maughan, et al., 1995). Researchers are finding that SSLPs possess a few advantages over traditional marker types. RFLPs are technically more complex, they require larger amounts of starting DNA, are more expensive, time-consuming, and have poorer genetic resolution. In a study by Bligh et al. (1999) dealing with genotyping cultivars of brown and white milled rice, SSLPs were demonstrated to possess a few characteristics more beneficial than AFLPs in the analysis. Once the initial design of primers for specific loci was complete, SSLPs were cheaper, faster, and more tolerant of poor quality DNA. Aside from genotype identification and quality assurance of commercial lines (Vaughan et al., 2001), SSLP application has been proven in gene and QTL evaluations (Moncada et al., 2001), marker-assisted breeding (Chen et al., 1997), and gene cloning (D. Blair, Cornell University, unpublished data; Chen et al., 1997; Yu et al., 1994).

Many resistance genes have been identified and sequenced in plants. Analysis of these sequences has uncovered conserved motifs, such as nucleotide binding site (NBS) and leucine rich repeat (LRR) regions in the DNA, that confer resistance to a broad spectrum of pathogens (Yu et al., 1996; Collins et al., 2001). Both are signature regions in the DNA that encode for domains known to be involved in the majority of plant pathogen resistance pathways (Collins et al., 2001). This knowledge can be extrapolated across species to identify analogous sequences, resistance gene analogues (RGAs), which include these signature motifs and may be involved with some type of disease resistance. The insights provided by RGAs create an opportunity that can greatly facilitate disease resistance gene identification and cloning. RGAs have already been developed in rice (Mago et al., 1999), tomato (Foolad et al., 2002), *Brassica napus* (Fourmann et al., 2001), grapevine (Donald et al., 2002), wheat and barley (Collins et al., 2001), and soybean (Yu et al., 1996).

It has already been illustrated that marker technology can be applied to RGAs, and several RGA sequences have been mapped in soybean. Data from several soybean studies support clustering of RGA families that map in close proximity to known resistance genes (Yu et al., 1996; Kanazin et al., 1996; Graham et al., 2000). This provides a framework for cloning genes through a "candidate" gene approach. For example, RGAs identified by Hayes et al (2000) and Yu et al (1996) were pivotal in the rapid high-resolution mapping and cloning of soybean potyvirus resistance genes (Gore et al., 2002; Hayes et al., in press).

In this study, two types of PCR-based markers were used, SSLPs (designed from ESTs) and RGAs (derived previously based on cloned R genes). The markers were screened for their ability to detect variant alleles in twelve soybean lines. This type of analysis will assess the level of variation demonstrated by SSLPs within soybean ESTs and soybean RGAs. Powerful markers should be able to detect a high level of polymorphism (variation) between soybean lines. The main goal was to determine the appropriateness of these types of markers as molecular tools. The ESTs chosen for the project were identified as being disease-related and the association of RGAs with resistance genes had already been demonstrated. The overall intentions of this study are to provide markers that can be used to pinpoint the location of resistance genes.

Materials and Methods

Genetic Material

Twelve representative soybean materials consisting of wild and cultivated plant introductions from diverse origins in the Far East, commercial cultivars and breeding lines, and parental lines of mapping populations were used for analysis (Table 1). Lines were chosen to represent as much variability in soybean as possible, giving our markers the best opportunity to distinguish any genetic differences. Gene sources were selected to cover origins spanning from Asia (China, Taiwan, and Korea) to several states in America (Virginia, Illinois, Maryland and Arkansas). Maturity classes ranging from III to VI and X were represented by our samples. Phenotypic traits are very different between the twelve lines, this includes disease resistance responses. PI96983 is known to contain resistance genes for soybean mosaic virus, peanut mottle virus, and phytopthora,

while Lee68 is known to be susceptible to all three of these diseases. The sample provides a good representation of the soybean genome and its gene pool.

Commercial cultivars and breeding lines were obtained from the soybean-breeding program at Virginia Tech. Dr. R. Nelson, United States Department of Agriculture, Agriculture Research Service (USDA-ARS), at the University of Illinois supplied the plant introductions. Dupont/Pioneer also provided four private PHP lines used for breeding and as mapping population parents. Young leaves were collected and bulked from plants of each line to make DNA for the analysis. Soybean genomic DNA was extracted from powdered freeze-dried tissue with CTAB extraction buffer, adhering to the protocol described previously by Saghai Maroof et al. (1984).

Source of Disease-Related ESTs

DuPont kindly supplied information about approximately 25,000 ESTs. The following method was employed by DuPont to generate the set of sequences. Soybean plants were inoculated with various pathogens and given time to respond. Several cDNA libraries were then created from specific organs of the plants, such as root, stem, and leaf, etc. The libraries were then scanned and all possible ESTs were identified and sequenced. Personnel from our lab working with DuPont/Pioneer scientists, then BLASTed the ESTs against sequences that are known to function in disease resistance pathways. The results produced 1218 ESTs that were homologous to the known disease-related sequences and used for SSLP marker construction. Sequence size of each EST ranged from 400 to 600 nucleotides.

SSLP Design

Each disease-related EST was BLASTed against a local database consisting of soybean ESTs from the National Center of Biotechnology Information (NCBI) database. NCBI is publicly available on the internet, and is a bioinformatics tool that organizes the abundance of genomic data being discovered daily. Homologous sequences produced by BLAST were scanned for SSR and insertion/deletion mutations that may produce polymorphism at that location in different soybean genomes (lines). LaserGene was used to design primers that amplify 150 to 300 base pair regions of the sequence which include these potentially polymorphic regions. The majority of the SSLPs were designed from SSR-containing regions. Although not as abundant, additional primers were designed from homologous ESTs that contained gaps (insertions/deletions) between their sequences. A total of 202 polymorphic sites were identified (from the set of 1218 ESTs) and used for marker construction. SSLP markers were numbered starting with 001, in order with the sequence that was used in primer synthesis. A few sequences contained more than 1 polymorphic site that was used in primer design. These markers were given the same number designated for that sequence, but distinguished by using capital letters at the end of the number. For example, an EST that contained two polymorphic sites had two different sets of primers designed. Both markers would be given the same number (SSLP000), but a capital letter would be added to separate the markers (SSLP000A, SSLP000B).

RGA Marker Design

Yu et al (1996) and Saghai Maroof (unpublished) have identified 22 RGAs in soybean. They appear to belong to a superfamily of candidate disease-resistance genes based on a conserved NBS motif. Mapping data has associated this motif with several known soybean-resistance genes (Yu et al., 1996; Hayes et al., 2000; Gore et al., 2002). The sequences for these RGAs were available in our lab for marker construction. Analysis for identification of polymorphic sites was not conducted. LaserGene was used to design primers as close to the beginning and end of the sequences as possible, leaving detection of polymorphism to chance.

Marker Analysis

SSLP and RGA screening procedures were identical to microsatellite analysis as described by Saghai Maroof et al. (1994) and Yu et al. (1994). Briefly, a 10 ul PCR reaction contained 50 ng of genomic DNA, 0.1 mM of each primer, 10x reaction buffer, 3 mM MgCl2, 200 mM each dATP, dGTP, dTTP, 5 mM dCTP, 1.0 U of AmpliTaq DNA polymerase (Perkin-Elmer Cetus, Norwalk, Ct.), and 0.05 uCi [ά-^{32}P] dCTP. The reaction was denatured at 94˚C for 3 minutes, followed by 32 cycles at 94˚C for 30 seconds, 47˚C for 30 seconds, and 68˚C for 1 minute, with a final extension step at 68˚C for 7 minutes. PCR products were denatured at 94˚C for 8 minutes after loading buffer (98% deionized formamide, 10 mM EDTA (pH 8.0), 0.25% xylene cyanol FF, 0.25% bromophenol blue) was added. Five microliters of each sample was loaded on a polyacrylamide denaturing gel and separated at 1500-V constant power in 1x TBE (Tris-Borate-EDTA) running buffer, using a DNA sequencing unit (Model STS-45, IBI, New-Haven, Ct.). Gels were immediately covered with plastic wrap and exposed to X-ray film.

Evaluation and Statistical Analysis

Each marker type (SSLP and RGA) was evaluated as a whole based on its ability to detect at least one variant in the set of representative soybean lines. A percentage of the number of markers that can detect polymorphism was calculated. Diversity level of each locus was also evaluated using a genetic diversity index ($1 - \Sigma p_i^2$), where p_i is the frequency of the i^{th} allele, as described by Yang et al (1994) and Nei (1973). Different alleles detected at each locus were labeled with letters of the alphabet. The smallest allele began with letter A, and allele size increased in alphabetical order (allele B larger than allele A, allele C larger than allele B, etc.).

The SSLP and RGA data were used to analyze genetic similarity among the soybean lines. Effective molecular markers are capable of separating plant lines into groups based on genes they share. Lines that are closer to the same origin and have common genes and characteristics should be grouped together. Genetic relatedness was assessed phenetically with the computer program POPGENE, version 1.32 (Yeh et al., 1997). The program calculated genetic identity and distances using Nei's statistics

(1972), and subjected them to cluster analysis by unweighted pair-group method analysis (UPGMA). This data was then used by POPGENE to draw a dendrogram that grouped related soybean lines.

Results

SSLP Polymorphism in the Total Sample

Polymorphic alleles differ in size and will create allelic ladders when visualized according to the marker analysis method described above (see Figure 1). Each step of the ladder represents a different size band and can be used to differentiate genotypes. Alleles were scored as previously described in the methods section and tabulated. The letter N was used to represent polymorphism detected by the absence of an allele (null allele). If multiple alleles were present in one line, they were scored as a variant allele and all corresponding letters were used for labeling. This system was implemented for the analysis of all 202 SSLPs designed.

An amplification product was produced by 187 of the 202 SSLP markers. Incompatibility that prevents primers from annealing to the soybean DNA will result in no amplification. It was observed that 66 of the markers that amplified (32% of the 202 initial SSLPs identified) detected at least two different genotypes in the twelve lines. In the analysis of the 12 lines using the 66 polymorphic SSLP markers, 213 variants were identified (Table 2). Markers SSLP012 and SSLP021 amplified two different regions of the genome allowing them to detect polymorphism for two different loci. A lowercase letter was used to distinguish the two marker loci, SSLP012a, SSLP012b, SSLP021a, and SSLP021b. The most variants detected per marker were 9, by SSLP022. SSLP020 and SSLP023 detected 8 variants, SSLP021b, SSLP053, and SSLP060 detected 7 variants, and SSLP009, SSLP009A, SSLP024, SSLP061, and SSLP090 detected 6 variants. On average, polymorphic markers were able to distinguish three genotypes (variants). 17% of the 66 SSLP markers detected a minimum of 6 variants, which allows them to differentiate at least half of the lines screened.

The diversity values calculated for each marker ranged from 0.152 to 0.861, with an average of 0.441 (Table 2). To gain a better understanding of how the majority of markers performed, their frequency was calculated according to four diversity classes: 0-0.25, 0.26-0.5, 0.51-0.75, and 0.76-1. Most markers, 47%, were grouped into the second class, 0.26-0.5. Only 23% were in the lower class, 0-0.25, and the higher diversity classes were almost even with 18% in 0.51-0.75 and 12% in 0.76-1 range.

SSLP technology is an improvement of microsatellite markers because it includes more types of sequence polymorphism that can be detected by markers. Microsatellites only focus on SSR regions in the genome, where SSLPs also include insertion or deletion polymorphisms that result in gaps. In our analysis, SSLP application provided additional markers for screening the twelve lines. The number of polymorphic loci identified that were useful in the diversity analysis was increased by the extra markers.

Comparison of RGA and SSLP Polymorphism

Both types of our markers were observed to have excellent amplification in the soybean genome, with primers designed from RGAs, at 95%, exhibiting a slightly higher rate of amplification success than SSLPs, at 92%. There has been concern about over-redundancy caused by the use of conserved sequences for primer annealing. The motifs of RGAs are repeated throughout the genome. It would be possible for the primers to anneal at many different regions on the DNA, causing numerous amplification products to be generated that could confound the identification of one useful marker. No discernable difference was observed between the clarity of amplification products for SSLPs and RGAs. Markers were easily identified and scored as monomorphic or polymorphic in the RGA analysis.

Polymorphic RGAs also distinguished three genotypes (variants) on average. However, RGAs were slightly lower (5%) in the number of markers that detected variant alleles, 6 out of 22. Although the maximum number of variants detected per marker was lower in RGAs, their average genetic diversity was 0.06 higher than the SSLPs. Not enough RGAs were used in the analysis to distribute them into diversity classes, as was done for the SSLPs.

Cluster Analysis

A good indicator of effective molecular markers is their ability to distinguish and separate plant lines that are related. The twelve lines used in our analysis provide a large range of genetic origins but have enough common characteristics, such as maturity class and cultivation for crop production, that should allow grouping. Genetic distances were computed and clustered by POPGENE. The resulting dendrogram (Figure 2) illustrated the ability of the markers to classify related genomes (lines).

PHP lines will not be discussed to keep their confidentiality in accordance with Dupont. A clear distinction was made between G. *max* and G. *soja* lines. The wild type soybeans (PI245331 and PI407162) were very distant from the other lines, and even demonstrated large separation between each other. The cultivars grouped relatively close to each other, with Peking being separated the farthest away. Between Peking and three of the cultivars were two closely related lines, V71-370 and PI96983, which are both in the maturity class V. The two cultivars that were most related originated in the United States, Maryland and Illinois, with the next closest line coming from Arkansas. The markers were able to group the lines based on origin and common traits which are a direct reflection of genetic similarities.

Genetic differences were also detected by the markers. PI96983 contains resistance genes for soybean mosaic virus, peanut mottle virus, and phytopthora, while Lee68 is clearly susceptible to all three of these diseases. The lines were placed in separate groups, indicating less relation between their genetic inheritances.

Discussion

The informative nature of ESTs and RGAs has provoked an interest in them as a source for molecular markers. Unlike random genomic sequence information, ESTs and

RGAs are specifically selected based on their involvement with genes of interest. Using sequences that are known to be functional and related to disease resistance pathways should facilitate the identification and tagging of resistance genes. First, it must be shown that PCR amplification can be achieved in the soybean genome from ESTs and RGAs. Most importantly, it must be determined that the markers are capable of detecting a high level of polymorphism. An SSLP analysis was chosen because it has emerged as a premier tool for rapid development of markers that are highly transferable and easily reproducible among labs. In this study, it is intended to show that the benefits of SSLP analysis also extend to ESTs. SSLP results can then be used to provide insights for the evaluation of RGA marker analysis through comparison.

Using ESTs as a source for markers has been demonstrated in rice (Wang et al., 2001), grapes (Scott et al., 2000), and wheat (Eujayl et al., 2002), but had not been explored in soybeans. PCR products were amplified by 92% of primers designed from ESTs, illustrating excellent primer recognition in the soybean genome. Molecular markers constructed from ESTs should also be useful in soybean.

Work by Cho et al (1998) and Moncada et al (2001) has already proven the highly polymorphic nature of SSLPs in rice. However, only non-coding regions of genomic sequences were used for analysis. The ability of SSLPs in soybean ESTs to detect polymorphism is clearly supported by the results of this study. The percentage of our SSLPs that detected polymorphism is slightly higher than the 25% observed in wheat EST-SSRs (Eujayl et al., 2002). In contrast, our rate of polymorphism was considerably lower than the 90% found by Moncada et al. (2001). However, this should not be surprising due to the high level of genetic diversity that rice is known to possess over most plant species.

SSLP application is considered to be an improvement over SSRs because it utilizes more sequence information to search for polymorphism. Thus, SSLPs should result in the development of more markers. In our study, additional markers have been identified for gene research in soybean that would have remained undiscovered had we limited are marker design to SSR analysis. The time and effort required to produce the additional insertion/deletion SSLPs was as efficient as for the other SSLPs (SSRs). Our findings demonstrate the benefits of SSLP technology.

The power of those SSLPs that were polymorphic was evident and encouraging. The high level of polymorphism strongly agrees with the highly polymorphic nature of microsatellites in soybean (Akkaya et al., 1995; Maughan et al., 1995). The diversity values calculated and reported in Table 2 can be used by other researchers to select the most powerful SSLPs. Markers with a higher value can distinguish a greater number of genotypes and will most likely be applicable to a greater number of mapping populations. Markers that are transferable across linkage maps, make it much easier to relate research from different studies. SSLP020, 021b, 022, 023, 053, and 060 would make excellent candidates.

Table 2 may also be useful for plant breeders wishing to quickly identify markers that might be useful for tagging a gene of interest. Markers not involved with the desired trait can be discarded by reviewing allele types on the table, saving time and effort that would otherwise be wasted on further testing. Markers detecting allele types that are identical between the line containing the desirable trait and the other lines that do not contain the trait should have nothing to do with the gene of interest. Thus, the marker

can be eliminated from the set. Working with a smaller set of markers should be much more effective for marker assisted selection breeding programs and gene cloning studies.

RGAs have already been successfully mapped in rice (Mago et al., 1999), soybean (Yu et al., 1996; Gore et al., 2002), barley (Collins et al., 2001), and their use is rapidly being expanded to many other crops. Yu et al. (1996) showed that RGAs could be mapped as RFLPs and amplified with primers in soybean. In the present study, using RGA sequences, we designed primers for PCR amplification and compared their activity with SSLPs.

The results of the analysis seem to be very similar for both sets of markers. It should be noted now that since such a few number of RGAs were tested (22), no statistical significance could be inferred from our results. However, the insights provided by any notable observations are still valuable. Considering that RGAs were not scanned for polymorphic regions in their sequences like the SSLPs, a rate of polymorphism as high as 27% is remarkable. Resistant genes have emerged over time due to mutations that alter into a needed function. Thus, our primers were scanning mutational hot spots. These regions increase the chances of producing a detectable polymorphism, which may have led to the success of the analysis.

We were also able to analyze the genetic similarities among the twelve soybean lines, further illustrating the usefulness of our markers. The G. *max* and G. *soja* were easily separated on our dendrogram, which is expected. It is well known that wild type and cultivated soybean are very diverse. The wild soybean line, PI245331, was most distantly related to the cultivars and has been reported as possibly being a distinct subspecies within the genus *Glycine* (Bernard et al., 1989; Maughan et al., 1996). Our observation supports the findings of these two studies. Peking, an ancestral line, was also clearly separated from the other cultivars by our markers. The results of our SSLP and RGA data are very similar to other markers that have been tested in soybean. Peking has also been separated from the majority of other cultivars by AFLPs (Maughan et al., 1996) and RFLPs (Skorupska et al., 1993). The potential of our markers seems to be clearly demonstrated by our results.

In addition, it seemed apparent that shared genes could also be identified. Those cultivars that originated in America and are drawing from the same gene pool were grouped the closest. Also, the more phenotypic traits that the lines shared, such as maturity class, the closer they were related by our analysis. At the same time, lines that differed according to resistance (PI96983) or susceptibility (Lee68) to soybean mosaic virus, peanut mottle virus, and phytopthora were related farther from each other and placed in separate groups. These results illustrate how desirable characteristics, such as disease resistance and high yield, can potentially be monitored and differentiated by our markers. The association of markers with disease resistance in plants can lead to the tagging, marker assisted selection, and cloning of resistance genes.

The information provided by our dendrogram can be used in a couple of ways to aid in genetic linkage map construction. Mapping populations can be designed to target specific traits, or to provide ample genetic variation that will increase the number of markers that can be mapped. For instance, a V71-370 cross with PI407162 would create ideal conditions for detecting polymorphic (variant) alleles. The two lines are distant according to our cluster analysis and have very different agronomic traits. PI407162 is a wild soybean that produces small seed, low in sucrose and yield. V71-370 is the exact

opposite, producing large, high sucrose content seed. This scenario indicates that a high level of genetic variation exists that may be related to genes of agronomic interest.

SSLPs have been used for genotyping, gene tagging, and linkage map construction in rice (Cho et al., 1998; Vaughan et al., 2001). It has been shown that this type of analysis is an improved molecular marker strategy that can easily be applied to soybean ESTs, which should facilitate gene tagging and gene cloning. The results of this report support similar conclusions between the RGA and SSLP analysis. These findings are promising for the future use of RGAs as molecular markers. Work by Yu et al. (1996), Hayes et al. (2000), Gore et al. (2002), who have all mapped RGAs in soybean, and Mago et al. (1999), who mapped RGAs in rice, help support this trend. This study confirms that EST and RGA PCR markers can be rapidly developed and very beneficial for genotyping in soybean. Marker assisted breeding programs and gene cloning research related to disease resistance in soybean are in need of powerful molecular tools like the ones reported in this study.

References

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

Bernard, R., Juvik, G., Nelson, R. 1989. USDA soybean germ plasm collection inventory. INTSOY Series Number 31, vol. 2.

Bligh, H.F.J., Blackhall, N.W., Edwards, K.J., and McClung, A.M. 1999. Using amplified fragment length polymorphisms and simple sequence length polymorphisms to identify cultivars of brown and white milled rice. Crop Sci. **39**: 1715-1721.

Chen, X., Temnykh, S., Xu, Y., Cho, Y.G., McCouch S.R. 1997. Development of a microsatellite framework map providing genome-wide coverage in rice (*Oryza sativa* L.). Theor Appl. Genet. **95**: 553-567.

Cho, Y.G., McCouch, S.R., Kuiper, M., Kang, M.-R., Pot, J., Groenen, J.T.M., and Eun, M.Y. 1998. Integrated map of AFLP, SSLP, and RFLP markers using a recombinant inbred population of rice (*Oryza sativa* L.). Theor. Appl. Genet. **97**: 370-380.

Clemente, T., LaVallee, B., Howe, A., Ward, D-C., Rozman, R., Hunter, P., Broyles, D., Kasten, D., and Hinchee, M. 2000. Progeny analysis of glyphosate selected transgenic soybeans derived from Agrobacterium-mediated transformation. Crop Sci. **40**: 797-803.

Collins, N., Park, R., Spielmeyer, W., Ellis, J., and Pryor, A.J. 2001. Resistance gene analogs in barley and their relationship to rust resistance genes. Genome **44**: 375-381.

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

Donald, T.M., Pellerone, F., Adam-Blondon, A.-F., Bouquet, A., Thomas, M.R., and Dry, L.B. 2002. Identification of resistance gene analogs linked to a powdery mildew resistance locus in grapevine. Theor. Appl. Genet. **104**: 610-618.

Eujayl, I., Sorrells, M.E., Baum, M., Wolters, P., and Powell, W. 2002. Isolation of EST-derived microsatellite markers for genotyping the A and B genomes of wheat. Theor. Appl. Genet. **104**: 399-407.

Foolad, M.R., Zhang, L.P., Khan, A.A., Nino-Liu, D., and Lin, G.Y. 2002. Identification of QTLs for early blight (*Alternaria solani*) resistance in tomato using backcross

populations of a *Lycopersicon esculentum x L. hirsutum* cross. Theor. Appl. Genet. **104**: 945-958.

Fourmann, M., Charlot, F., Froger, N., Delourme, R., and Brunel, D. 2001. Expression, mapping, and genetic variability of *Brassica napus* disease resistance gene analogues. Genome **44**: 1083-1099.

Gore, M.A., Hayes, A.J., Jeong, S.C., Yue Y.G., Buss, G.R. and Saghia Maroof, M.A. 2002. Mapping tightly linked genes controlling potyvirus infection at the *Rsv1* and *Rpv1* region in soybean. Genome. **45**: 592-599.

Graham, M.A., Marek, L.F., Lohnes, D., Cregan, P., and Shoemaker, R.C. 2000. Expression and genome organization of resistance gene analogs in soybean. Genome **43**: 86-93.

Hayes, A.J., Yue, Y.G., and Saghai Maroof, M.A. 2000. Expression of two soybean resistance gene candidates shows divergence of paralogous single-copy genes. Theor. Appl. Genet. **101**: 789-795.

Holton, T.A., Christopher, J.T., McClure, L., Harker, N., and Henry, R.J. 2002. Identification and mapping of polymorphic SSR markers from expressed gene sequences of barley and wheat. Molecular Breeding **9**: 63-71.

Kanazin, V., Marek, L.F., and Shoemaker, R.C. 1996. Resistance gene analogs are conserved and clustered in soybean. PNAS **93**: 11746-11750.

Liu, K. 1997. Soybeans: Chemistry, Technology, and Utilization. Chapman and Hall. New York, New York.

Mago, R., Nair, S., and Mohan, M. 1999. Resistance gene analogues from rice: cloning, sequencing and mapping. Theor. Appl. Genet. **99**: 50-57.

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

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

Moncada, P., Martinez, C.P., Borrero, J., Chatel, M., Gauch, H., Guimaraes, E., Tohme, J., McCouch, S.R. 2001. Quantitative trait loci for yield and yield components in an Oryza sativa x Oryza BC_2F_2 population evaluated in an upland environment. Theor. Appl. Genet. **102**: 41-52.

Nei, M. 1972. Genetic distance between populations. Am. Nat. **106**: 283-292.

Nei, M. 1973. Analysis of gene diversity in subdivided populations. PNAS **70**: 3321-3323.

Riley, P. 1999. USDA expects record soybean supply. Inform, **10**: 503-506.

Saghai Maroof, M.A., Soliman, K.M., Jorgenson, R.A., and Allard, R.W. 1984. Ribosomal DNA spacer length polymorphism in barley: Mendelian inheritance, chromosomal location, and population dynamics. PNAS **81**: 8014-8018.

Saghai Maroof, M.A., Biyashev, R.M., Yang, G.P., Zhang, Q., and Allard, R.W. 1994. Extraordinarily polymorphic microsatellite DNA in barley: Species diversity, chromosomal locations, and population dynamics. PNAS **91**: 5466-5470.

Scott, K.D., Eggler, P., Seaton, G., Rossetto, M., Ablett, E.M., Lee, L.S., and Henry, R.J. 2000. Analysis of SSRs derived from grape ESTs. Theor. Appl. Genet. **100**: 723-726.

Shoemaker, R., Keim, P., Vodkin, L., Retzel, E., Clifton, S.W., Waterston, R., Smoller, D., Coryell, V., Khanna, A., Erpelding J. et al. 2002. A Compilation of soybean ESTs: generation and analysis. Genome **45**: 329-338.

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

Sterky, F., Regan, S., Karlsson, J., Hertzberg, M., Rohde, A., Holmberg, A., Amini, B., Bhalerao, R., Larsson, M., Villarroel, R., Van Montagu, M., Sandberg, G., Olsson, O., Teeri, T.T., Boerjan, W., Gustafsson, P., Uhlen, M., Sundberg, B., and Lundeberg, J. 1998. Gene discovery in the wood-forming tissues of popular: analysis of 5,692 expressed sequence tags. PNAS **95**: 13330-13335.

Vaughan, L., Ottis, B., Prazak-Harvey, A., Sneller, C., Chandler, J., and Park, W. 2001. Is all red rice found in commercial rice really *Oryza sativa*? Weed Science **49**: 468-476.

Wang, Z., Taramino, G., Yang, D., Liu, G., Tingey, S.V., Miao, G.H., and Wang, G.L. 2001. Rice ESTs with disease-resistance gene or defense-response gene-like sequences mapped to regions containing major resistance genes or QTLs. Mol. Genet. Genomics **265**: 302-310.

Yang, G.P., Saghai Maroof, M.A., Xu, C.G., Zhang, Q., and Biyashev, R.M. 1994. Comparative analysis of microsatellite DNA polymorphism in landraces and cultivars of rice. Mol. Gen. Genet. **245**: 187-194.

Yeh, Francis C., Yang, R-C., Boyle, Timothy, B.J., Ye, Z-H., and MAO, Judy X. 1997. POPGENE, the user-friendly shareware for population genetic analysis.

Molecular Biology and Biotechnology Centre, University of Alberta, Canada.

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

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

Table 1. Twelve Soybean Lines Used for Diversity Analysis with SSLPs and RGAs

Name	Species	Origin	Maturity	Type[1]
Lee68*	G.*max*	Arkansas	VI	C
Peking	G.*max*	China	IV	C
Williams*	G.*max*	Illinois	III	C
Wye	G.*max*	Maryland	IV	C
V71-370*	G.*max*	Virginia	V	C
PI96983*	G.*max*	Korea	V	C
PI245331	G.*soja*	Taiwan	X	W
PI407162*	G.*soja*	Korea	IV	W
PHP1	-	-	-	D
PHP2	-	-	-	D
PHP3	-	-	-	D
PHP4	-	-	-	D

* Lines that have been used to develop a mapping population. (PI) indicates plant introduction lines.
[1] C- cultivars W- wild
D-parental lines provided by Dupont (information is confidential)

Table 2. Genetic Diversity and # of variants for 66 SSLPs and 6 RGAs Detecting Two or More Variants.

EST ID	Wye	Williams	PI96983	Lee68	PI407162	V71-370	PI245331	Peking	PHP1	PHP2	PHP3	PHP4	Div Ind.[1]	# variants
SSLP002	B	B	B	B	A	B	B	B	B	B	B	B	0.152	2
SSLP004	D	D	D	A	D	D	C	D	D	B	D	A	0.513	4
SSLP006	A	A	A	N	N	N	A	N	A	N	A	A	0.486	2
SSLP009	D	D	E	A	-	A	A	B	D	C	AD	-	0.780	6
SSLP009A[2]	D	D	E	A	D	A	A	B	-	C	AD	D	0.760	6
SSLP012a[3]	-	A	A	-	A	A	B	B	A	A	A	-	0.345	2
SSLP012b	-	A	D	-	D	D	C	E	B	D	B	-	0.716	5
SSLP013	B	B	C	B	N	B	D	A	B	B	-	B	0.562	4
SSLP015	B	B	A	B	A	A	A	B	B	A	AB	A	0.569	3
SSLP016	B	D	B	D	E	B	C	C	B	A	D	B	0.722	5
SSLP017	B	B	B	B	C	B	B	B	B	-	B	A	0.314	3
SSLP018	A	A	A	A	B	A	A	A	A	A	A	A	0.152	2
SSLP020	F	E	E	G	B	N	A	C	D	N	C	C	0.847	8
SSLP021a	A	B	A	A	B	A	B	A	B	B	B	B	0.486	2
SSLP021b	B	C	A	D	N	A	E	N	C	F	C	A	0.819	7
SSLP022	H	H	F	H	A	F	N	C	B	D	G	E	0.861	9
SSLP023	F	GHI	B	C	E	F	F	D	H	A	H	C	0.847	8
SSLP024	E	F	A	E	B	E	E	E	D	C	-	-	0.700	6
SSLP026	ABC	ABC	ABC	ABC	-	ABC	AC	AC	ABC	ABC	ABC	-	0.320	2
SSLP028	A	A	A	A	A	N	A	A	A	A	A	A	0.152	2
SSLP035	A	B	A	B	A	A	A	A	A	A	A	A	0.277	2
SSLP039	A	B	A	B	A	B	B	A	B	B	A	B	0.486	2
SSLP048	B	B	B	B	A	B	B	A	B	C	C	B	0.500	3
SSLP053	D	F	G	A	B	A	C	E	F	G	F	A	0.819	7
SSLP054	B	B	B	A	-	B	A	A	B	B	A	B	0.462	2
SSLP057	B	B	B	BC	B	C	B	B	B	A	B	B	0.416	4
SSLP058	N	N	N	N	N	N	A	N	N	N	N	N	0.152	2
SSLP060	F	F	B	F	C	A	D	A	A	C	AF	E	0.819	7
SSLP061	C	C	E	D	A	D	B	C	C	C	CD	C	0.694	6
SSLP070	A	A	A	A	A	A	B	A	A	A	A	A	0.152	2
SSLP079	A	A	A	A	B	A	A	A	A	A	A	A	0.152	2
SSLP083	A	A	A	A	A	C	A	A	B	A	B	A	0.402	3
SSLP084	A	B	A	A	A	-	B	A	B	A	B	A	0.462	2
SSLP090	D	A	A	B	B	C	E	D	A	A	F	A	0.750	6
SSLP091	A	A	A	A	B	A	B	A	-	A	A	-	0.320	2
SSLP093	C	C	A	C	A	A	A	B	C	C	C	C	0.541	3
SSLP094	A	A	A	A	A	B	A	A	A	A	A	A	0.152	2
SSLP100	A	A	A	A	C	A	B	A	A	A	A	A	0.291	3
SSLP102	A	A	A	A	D	A	D	B	C	B	A	A	0.597	4
SSLP107	B	B	B	B	B	B	B	B	B	A	B	B	0.152	2
SSLP112	B	B	B	B	B	B	A	B	B	B	B	B	0.152	2
SSLP113	A	A	A	A	A	A	B	A	A	A	A	A	0.152	2
SSLP114	B	B	B	B	C	B	A	B	B	B	B	B	0.291	3
SSLP116	A	A	A	A	B	A	A	A	A	A	A	A	0.152	2
SSLP117	A	AB	-	B	B	A	B	A	A	A	A	B	0.562	3
SSLP118	A	A	A	A	A	A	A	A	B	A	B	A	0.277	2
SSLP121	A	A	B	A	B	B	B	B	A	B	A	A	0.500	2

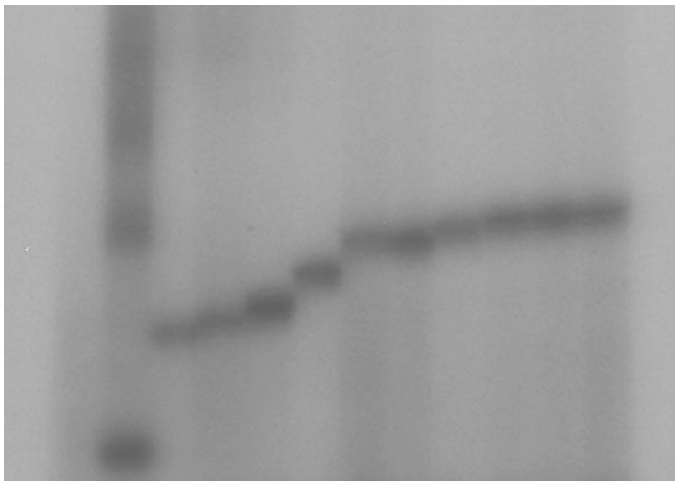
Table 2 continued, Genetic Diversity and # of variants for 66 SSLPs and 6 RGAs Detecting Two or More Variants.

EST ID	Wye	Williams	PI96983	Lee68	PI407162	V71-370	PI245331	Peking	PHP1	PHP2	PHP3	PHP4	Div Ind.	# variants
SSLP122	B	B	B	B	B	B	A	A	B	A	C	B	0.486	3
SSLP124	A	A	B	A	B	A	B	B	A	A	AB	A	0.541	3
SSLP125	A	A	A	B	B	B	-	B	B	A	A	A	0.495	2
SSLP130	N	N	N	N	N	N	A	N	N	A	B	N	0.402	3
SSLP132	A	A	A	A	A	A	B	A	A	A	A	A	0.152	2
SSLP140	A	A	A	A	N	A	N	A	A	A	A	A	0.277	2
SSLP142	A	A	B	-	A	B	A	-	A	A	B	A	0.420	2
SSLP143	B	B	B	-	A	B	B	B	B	A	B	B	0.297	2
SSLP145	A	A	A	-	A	A	A	A	A	B	A	A	0.165	2
SSLP147	B	B	B	-	B	AB	N	A	A	A	A	A	0.644	3
SSLP153	A	B	B	B	A	A	B	B	B	A	B	B	0.444	2
SSLP160	-	A	A	B	A	A	B	A	B	B	-	A	0.480	2
SSLP161	A	-	B	B	A	B	A	B	B	-	B	B	0.420	2
SSLP162	B	B	B	B	B	B	A	B	B	B	B	B	0.152	2
SSLP163	-	A	A	A	B	A	B	A	A	A	A	A	0.297	2
SSLP181	A	A	A	A	B	A	-	A	-	-	-	-	0.244	2
SSLP185	A	A	A	-	A	A	-	-	A	N	A	N	0.345	2
SSLP193	C	C	C	-	A	C	B	-	B	C	B	B	0.580	3
SSLP197	N	-	A	N	A	N	A	-	A	N	A	A	0.480	2
SSLP Ave[4]													0.441	3
SSLP Total[5]														213
RGA18	AB	B	B	-	A	B	-	B	AB	B	AB	B	0.540	3
RGA19a	N	AC	A	-	AB	A	N	A	N	C	N	-	0.720	5
RGA19b	B	B	A	-	A	B	B	A	B	A	B	-	0.480	2
RGA43	N	N	N	-	N	N	N	N	A	N	N	A	0.297	2
RGA57	B	BC	B	-	BC	B	ABC	AB	B	B	B	B	0.545	4
RGA307	B	B	A	-	A	B	B	A	B	A	B	B	0.462	2
RGA Ave.													0.507	3
RGA Total														18
Total[6]													0.451	231

Note: See materials and methods for detailed scoring methodology; letters represent allele size (variants). The detection of multiple variants (heterozygosity) in a parental line, indicated by labeling with more than one letter, was assessed as one variant.

(N) represents absence of allele (null). (-) indicates missing data.

[1]Diversity Index (Nei, 1973) used to assess markers ability to detect variation at that locus.
[2]Capital letters, in column one, differentiate markers (primer pairs) designed from the same EST.
[3]Lowercase letters, in column one, differentiate separate locations (loci) in the genome that were amplified by the same marker.
[4]Average values (diversity index and # of alleles) for the 66 SSLPs
[5]Total number of variants detected by SSLPs
[6]Values calculated for combination of 66 SSLPs and 6 RGAs. Diversity index is an average and # of variants is the total.



N A B C D E F F G H H H

Figure 1 Allelic ladder from diversity analysis of SSLP022, DNA from each of the twelve soybean lines was radioactively labeled, loaded into separate lanes, and separated on a polyacrylamide gel. The bands were then visualized with X-ray film. Nine variant alleles were detected as indicated by letter A-H and N (null allele) below each band. The parents from left to right are: PI245331, PI407162, PHP1, Peking, PHP2, PHP4, V71-370, PI96983, PHP3, Wye, Williams, and Lee68.

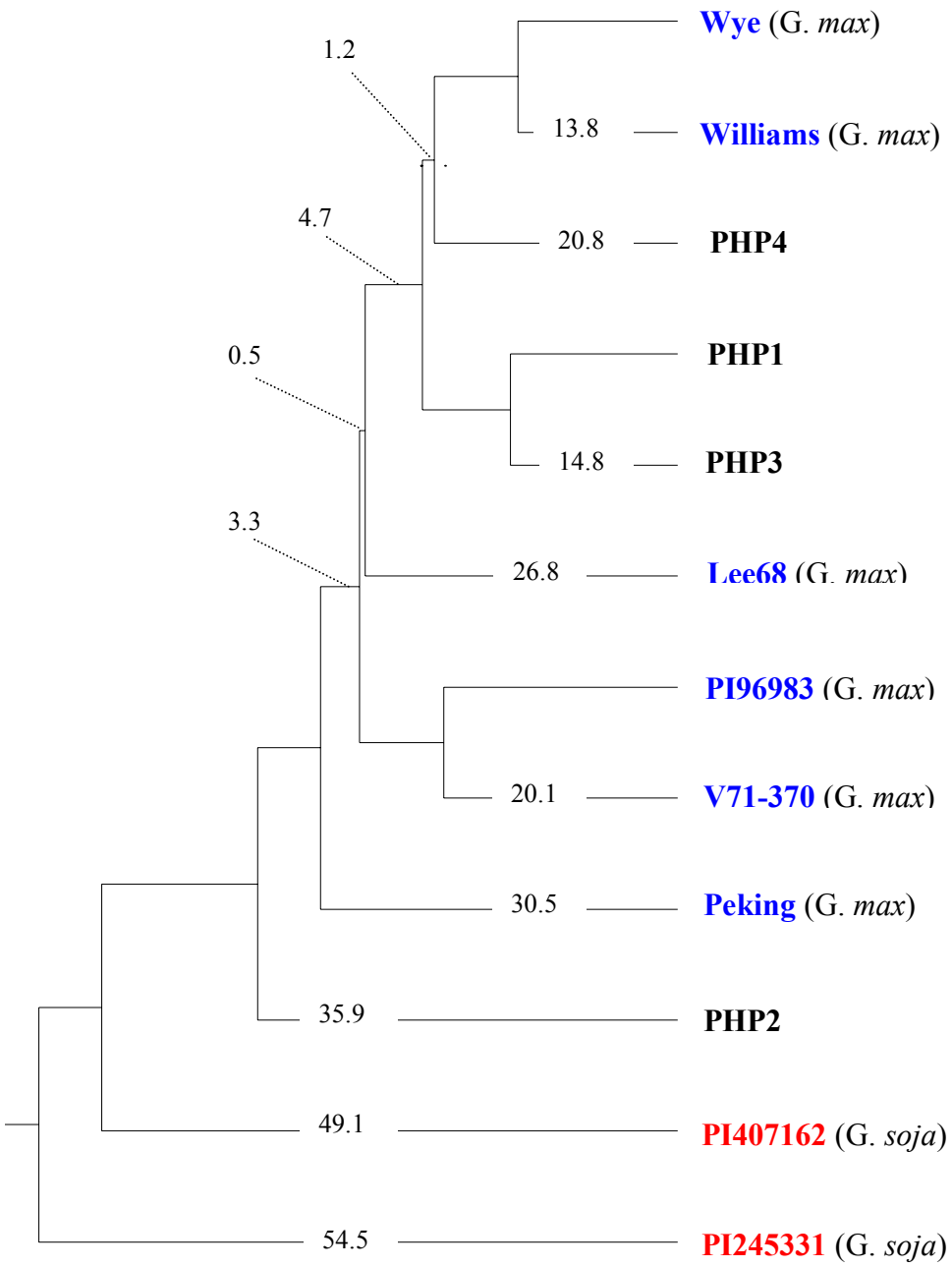


Figure 2 Dendrogram assessing genetic similarity of 12 soybean lines using SSLP and RGA marker data from the diversity analysis. Branch distances were obtained by cluster analysis (UPGMA) of Nei's (1972) measures of genetic distances. G. *max* is clearly grouped separately from the G. *soja*, as indicated by the long branch distances between them.

Chapter III. **Molecular Mapping of Disease-related ESTs and RGAs in Soybean.**

Abstract

Soybean is host to one of the widest range of infectious pathogens of all crops. The economic and social value of soybean makes the development of disease resistant cultivars crucial. Molecular markers are the premier tool to accomplish this task. In this study, 48 disease-related ESTs, five RGAs and 150 published molecular markers were used to construct a soybean linkage map. The published markers are known to cover regions that represent all 20 chromosomes in soybean. A population of 114 recombinant inbred lines (RILs) from a cross between an agronomic cultivar, V71-370, and a wild line, PI407162 that demonstrates much phenotypic difference was used. Six ESTs and four RGAs mapped to regions that contain known resistance genes or quantitative trait loci (QTLs) for various diseases. All RGAs were linked near one or more ESTs, which indicate that the markers may be parts of functional genes. Molecular linkage groups (MLGs) D1a and L possessed the highest concentration of mapped ESTs, each with 6. Several QTLs for soybean cyst nematode (SCN) have previously been identified on chromosome D1a, but no disease resistance gene has yet been discovered on chromosome L. The clustering of ESTs on linkage group L may indicate the location of new resistance genes. Several markers were mapped to regions previously known to be resistance loci, and thus may be useful for gene cloning. SSLP090 mapped to the chromosomal region containing *Rsv4*, a gene that provides resistance to all known strains of soybean mosaic virus (SMV). SSLP039 and RGA43 mapped to a region on the bottom of chromosome J near a cluster of disease resistance loci. This study has created a framework map for identifying the location of resistance genes, discovering novel resistance genes, and understanding genome organization of resistance pathways. The effectiveness of using EST and RGA markers for accelerating the candidate gene approach has been demonstrated by this study.

Introduction

Soybean is plagued by one of the broadest ranges of infectious diseases in crops, with over 100 pathogens infecting it (Sinclair et al., 1999). This negative attribute has resulted in extensive study of disease resistance genes in soybean. It has become an important task to identify all disease-related genes for manipulation in crop improvement. Molecular markers and their various applications have become pivotal in the advancement of this goal.

The primary use of molecular markers is the construction of genetic linkage maps. Molecular maps enhance the ability of plant geneticists to improve crops through marker assisted selection, to identify genes via map-based cloning, and to develop improved transgenic plants. The key is to develop markers that are transferable to existing maps, optimizing the information obtained from each marker. Linkage maps with different purposes can be integrated to saturate important regions of the genome. The work of Cregan et al (1999) led to the development of a globally edifying linkage map in soybean that was constructed with easily transferable SSR markers. Reproducible markers allowed the discoveries of independent studies to be related by a common set of markers. As a result of integral maps, soybean has experienced improvements from marker assisted selection for SMV resistance (communication with Dr. Buss and Dr. Tolin, Virginia Tech 2003), the identification of a novel disease resistance gene (Hayes et al., 2000b), the high-resolution tagging and cloning of potyvirus resistance genes (Gore et al., 2002), and the transformation of a putative SMV resistance gene (communication with Dr. Saghai Maroof, Virginia Tech 2003).

The progress of molecular markers has always directed the improvement and utility of genetic maps. In the early nineties, the low level of genetic polymorphism generated by RFLPs in soybean posed serious problems for the application of linkage maps (Shoemaker et al., 1996). The science was in need of techniques designed to detect a higher degree of genetic diversity. The work of Akkaya et al. (1992) and Rongwen et al. (1995) demonstrated that SSRs could be used to detect an increased level of polymorphism in soybean. By manipulating the strategy of marker design, the pitfalls experienced in linkage map construction were overcome. The expanding and diverse markets that have come about for soybean make it vital to continue upgrading molecular markers used in map construction. The embellishment of marker design is the only way to proliferate the use of molecular maps. The goal is to construct a marker that offers maximum information and can be used by other researchers.

ESTs represent functional genes and are a tremendous improvement of genomic sequence information that is primarily composed of repetitive DNA and non-coding regions. It is possible to design markers within the region of a gene and when the marker is mapped the precise location of the gene would be revealed. ESTs homologous to possible genes can be targeted for genetic mapping, and mapped ESTs may be found to directly affect a trait of interest. In addition, ESTs are more likely to be compatible with linkage maps across pedigrees and species because coding regions of DNA are generally highly conserved regions. These characteristics are not found in traditional markers and will allow researchers to revolutionize their approach at the molecular level. A "functional" marker is at hand that can greatly facilitate the unraveling of biological pathways and interaction of genes. In a paper by Thibaud-Nissen et al. (2003), the stages

and development of somatic embryogenesis in soybean were characterized by differentiating ESTs. Markers can now be designed from these genes and genome organization can be associated with the pathways defined by Thibaud-Nissen et al. In effect, the entire biological process has become subject to molecular marker applications.

Public databases are increasing in the number of soybean ESTs daily, making the sequences an excellent source to scan for polymorphism. Using EST information, markers have been designed and used to detect genetic variability in wheat, barley, rice, grape, and soybean (Holton et al., 2002; Wang et al., 2001; Scott et al., 2000; see Chapter II). Wang et al (2001) have also demonstrated the resourcefulness of ESTs when they mapped defense-response gene-like sequences in rice. Clustering of the ESTs near regions associated with disease resistance QTLs was observed. These are very promising results for the utility of EST markers in gene mapping.

Resistance gene analogues (RGAs) are a second type of sequence information with improved quality over random DNA markers. They are generated directly from conserved motifs that have been identified in disease resistance genes. This characteristic gives them transferability equivalent to or better than that present in ESTs. NBS and LRR motifs have been used to identify RGAs in *Brassica napus*, grapevine, barley, rice and soybean (Fourmann et al., 2001; Donald et al., 2002; Collins et al., 2001; Mago et al., 1999; Yu et al., 1996). There has been considerable effort to study RGAs in soybean. The analogues have been observed to be conserved and clustered (Kanazin et al., 1996), to correspond to expressed sequences (Graham et al., 2000), and associated with known resistance QTLs and genes based on high sequence similarity (Hayes et al., 2000a; Penuela et al., 2002). These are superb features for a candidate gene approach to identify new resistance genes.

The main objective of this study was to integrate disease-related ESTs and RGAs into an existing soybean linkage map. The initial map was constructed with published microsatellite markers that have been associated with the locations of many known disease resistance genes. These published resistance genes can be superimposed on the final map to determine their relationship with new EST and RGA markers, such as those from this study. The benefit and resourcefulness of the related markers for gene identification and future manipulation can be assessed.

Materials and Methods

Mapping Population

A total of 114 RILs were developed from the interspecific hybridization between an adapted, large-seeded (24 g/100 seeds), high sucrose (8.3%) *Glycine max* breeding line (V71-370), and a small seeded (1.8 g/100 seeds), low sucrose (1.6%) *Glycine soja* plant introduction (PI407162). The extensive phenotypic differences are expected to provide many genetic variations between the two genotypes that can be detected by our markers. The goal of choosing these two differential lines was to maximize the number of markers that could be mapped.

The RILs were developed using a modified single-seed descent method. Four seeds were chosen at random from each plant in each generation and bulked, starting with the F_2 that was grown in the field at Eastern Virginia Agriculture Research and Extension

Center, near Warsaw, Virginia. Lines were carried forward from the F_3 to the F_8 generation, allowing plants to self-pollinate in the greenhouse at Virginia Tech. A single seed was used in the generation F_5 through F_8 in the greenhouse. Bulked seed from each F_8 plant was grown in the field in individual rows at Warsaw, Virginia, in 1998. Young leaves were then taken from each plant to be used for DNA extraction. These RILS were previously genotyped with a set of RFLPs and SSRs (Cicek, 1997.)

DNA Extraction

DNA from each parent and RIL was isolated (Cicek, 1997) from their leaves according to the protocol described by Saghai Maroof et al. (1984). Freeze-dried leaflet tissue (0.75 g, dry weight) was ground with a mechanical mill, dispersed in 15 ml of CTAB extraction buffer (50 mM Tris, 0.7 M NaCl, 10 mM EDTA, 1% hexadecyltrimethylammonium bromide, 0.1% 2-mercaptoethanol), and incubated at 65˚C for 60 minutes in a shaker bath. Ten ml of chloroform/octanol (24:1) were added. The solution was mixed by inversion and then centrifuged at 3200 rpm at 4˚C for 15minutes. The aqueous phase was transferred to a new tube. Two third volume of isopropanol was added and mixed by inversion. The precipitated pellet of DNA was hooked with a glass rod and transferred to a glass tube containing 20 ml 76% EtOH / 10 mM NH4Ac. After overnight washing in the EtOH solution, the DNA pellet was air dried and then dissolved in 10 mM NH4OAc / 0.25 mM EDTA. A fluorometer reading was taken for each F_8 DNA sample in order to measure the DNA concentration. Uncut DNA was run on 0.8% agarose gel to check its quality.

Sequencing of Disease-Related ESTs and Marker Design

The generation of ESTs that were determined to be homologous with genes associated with stress and disease response in soybean is described in Chapter II. The sequences were scanned for SSLPs to be used in marker synthesis; these procedures are also found in Chapter II. All markers that represent the disease-related ESTs will be referred to as SSLPs in this paper, with the exception of six SSRs designed by Cregan et al (1999), Sat062, Sat085, Sat105, Satt368, Satt398, and Satt407. We identified polymorphism in six disease-related ESTs that corresponded to sequences of these six published primers. To avoid confusion, these primers were ordered and used to represent the ESTs. These markers were also beneficial in defining the skeletal map.

SSLP analysis was chosen because it has repeatedly proven its ability to detect a high level of polymorphism (Moncada et al., 2001; Bligh et al., 1999; Chen et al., 1997; Maughan, et al., 1995). It includes the benefits of SSR analysis, as well as encompassing a broader spectrum of polymorphism (e.g., additions/deletions). The likelihood of identifying a polymorphic site is increased. SSLPs can also easily be integrated with other types of markers, as demonstrated by Cho et al (1998).

RGAs have been a good source for the candidate gene approach in disease resistance studies. It would be very beneficial to identify the chromosomal positions of as many RGAs as possible. This will facilitate the identification and cloning of many novel resistance genes. Thus, markers were designed for all possible soybean RGAs identified in our lab. Chapter II explains the development of these markers.

Marker Analysis

From the results of Chapter II, all SSLP markers that identified different genotypes between the parents, V71-370 and PI407162, used to develop the mapping population were determined. Furthermore, published SSR primers detecting polymorphism between the parents were selected, approximately 20 cM apart, to span each chromosomal linkage group. All markers were applied to the 114 RIL mapping population using PCR analysis described in Chapter II. For certain markers, the PCR conditions were manipulated to improve the distinctness of amplification products. In general, increasing the number of cycles was sufficient to increase band intensity and clearness. One marker required a lower annealing temperature. All conditions that were manipulated for each primer pair are reported in Table 1. For a given marker, individuals that inherited the allele of V71-370 were scored as A, and individuals that inherited the allele of PI407162 were scored as B. The inheritance pattern of each marker was recorded and used for recombination analysis in map construction.

Map Construction

Mapmaker 3.0 was used to group and order the genetic loci in the study. The published primers represented a skeletal map of each chromosome that was used to identify the positions of the unknown SSLPs and RGAs. Marker loci were grouped at LOD 3.0 (logarithm to the base 10 of likelihood odds ratio) with a maximum Haldane distance of 50 centiMorgans. Previous studies in our lab (Hayes et al., 2000; Gore et al. 2002; unpublished data) have involved the mapping of disease resistance loci using microsatellite markers in the same population as this study. All markers related to these loci were added to the above linkage analysis. Publically available SSR markers were added to the representative linkage groups throughout the project in an effort to close all gaps.

Many disease resistance genes have been mapped in soybean using microsatellites markers. The literature was reviewed to associate the position of these resistance loci to markers on our map. The Soybase website was an instrumental resource in tracking down references that identify resistance genes and their relation to several microsatellite markers (Shoemaker 2003). The position of tightly linked SSRs could then easily be related to many different types of markers using Cregan's published map (1999).

Identifying EST Function

All mapped SSLPs were BLASTed against a local database containing soybean ESTs and their putative function. If the identity of several homologous sequences were high and they agreed on putative function, then the SSLPs were also assigned the function. The nucleotide sequences of the BLAST matches were then converted to protein sequences using NCBI's BLASTx program (2003). There are many Bioinformatics tools that can be used to analyze protein sequences for signature profiles or domains. Many of these sequence characteristics can be categorized to specific functions. The University of Minnesota (2002) has developed an easily accessible internet program, "Metafam", that allows several of these tools to be harnessed at one

time. The protein sequences we obtained were entered into "Metafam" and analyzed by the following programs: BLOCKS version 01 (Fred Hutchinson Cancer Research Institute, 2001), Prosite version 17.2 (North Carolina Supercomputing Center, 2002), and prints version 32 (Washington University, 2003). Descriptions of domains or profiles identified are provided by all of the programs and can be combined to ascertain the best understanding of EST function.

Phylogenetic Analysis

ESTs that were assigned putative function were subjected to phylogenetic analysis using Biology Workbench (2003), provided by the San Diego Supercomputing Center (SDSC). To maintain privacy of Dupont's ESTs, homologous sequences were identified in GenBank and used to create a multiple sequence alignment with CLUSTALW (Thompson et al., 1999) default settings. The sequence alignment was imported into DRAWGRAM (Felsenstein, 1993) to construct a rooted phylogenetic tree; no branch length setting was applied. Default settings were used for all other commands.

Results

Skeletal Map

Mapping data were collected for 150 published SSRs; this number includes the 6 SSRs used to represent our ESTs. In the linkage analysis, 143 of these SSRs mapped to a linkage group. The unlinked markers were located near the tips of chromosomes and have distances that are far from the closest markers included in the analysis. Linkage groups were identified for all 20 chromosomes and are reported on Figure 1; however, some groups were separated by gaps producing internal groupings. It can be seen from Figure 1 that A2, B1, B2, D1a, and J were separated in two groups, while C1 and E were separated in three groups. Even though some linkage groups contained gaps, the overall span of each chromosome was covered by the markers mapped. A good basis is provided for the detection of SSLPs or RGAs linked to any chromosome. Thus, the integration of these new types of markers should be possible. The linkage distances covered by the skeletal map should be within the detectable limits of Mapmaker's linkage analysis.

Superimposing Disease Resistance Genes on the Skeletal Map

Cregan et al (1999) have developed a large set of easily reproducible SSR markers that have been mapped in three different populations. The locations of the markers are understood and this has allowed many researchers to harness the power of the SSRs for targeting resistance genes. Published data illustrating these results can be used to infer the correlation between our markers and the resistance genes through the comparison of SSRs included on our map.

Regions associated with the defense of the following 12 diseases were identified on our map, Bacterial Blight, Bacterial Pustule, Bud Blight, Brown Stem Rot, Corn Earworm, Frogeye Leaf Spot, Javanese Root-knot nematode (JRN), Peanut Mottle Virus, Peanut Root-Knot Nematode, Phytopthora Stem and Root Rot, Powdery Mildew,

Nodulation, Sclerotinia Stem Rot (SR), Southern Root-Knot Nematode (SKN), Soybean Cyst Nematode (SCN), Soybean Mosaic Virus (SMV), and Sudden Death Syndrome. A total of 16 resistance loci and 18 QTLs were identified in the analysis and have been superimposed on our linkage map in Figure 1. The associations uncovered by our literature review are also recorded in Table 2.

Distribution of SSLPs and RGAs in the Soybean Genome

Mapping data were collected for 55 SSLPs (including the 6 published SSRs) and 5 RGAs. Mapmaker assigned all 5 RGAs and 48 of the SSLPs to various linkage groups. All mapping results are illustrated on Figure 1. The majority of SSLPs were randomly distributed among linkage groups (Table 3); A1 and N were the only linkage groups to which no SSLPs mapped. The highest concentration of SSLPs were on linkage groups L and D1a, each having 6 located on them. C1, C2, and I all had 4 ESTs mapped to them, with the rest of the linkage groups averaging 2 ESTs. A cluster of three SSLPs were observed near the tip of linkage group L. Three SSLPs were linked to each other, but their MLG could not be identified because no published markers were mapped with them (Figure 1). Our linkage analysis, left seven SSLPs (including the three linked SSLPs without an identified MLG) and seven published markers unmapped. The unlinked published markers are located near the tips of chromosomes and have distances that are far from the markers mapped in their respective linkage group.

48 SSLPs (including the six published markers that represent SSLPs) and five RGAs were linked to published primers, allowing their integration into the skeletal map (Table 3). By associating our markers with the published markers used to superimpose disease resistance genes on the skeletal map (Figure 1 and Table 2), a total of 13 SSLPs and four RGAs were related to resistance loci (Table 3). All four RGAs and six of the SSLPs were determined to be in close proximity to several of the loci. SSLP008, SSLP017, and SSLP053 are near QTLs that confer resistance to *sclerotinia* stem rot (MLG K), soybean cyst nematode (MLG G), and cornear worm (MLG M), respectively. Three of the RGAs, RGA19a, RGA19b, and RGA307, were also mapped near QTLs, all conferring resistance to soybean cyst nematode (MLG M and E). RGA43 is near a region on the bottom tip of MLG J that is clustered with seven resistance genes for brown stem rot, frogeye leaf spot, powdery mildew, phytopthora, and nodulation. Interestingly, a SSLP also mapped with this RGA at that region on MLG J (SSLP039 and RGA43; Fig. 1, Table 3). This observation may indicate functionality of the RGA as one of the seven resistance genes at that location. Figure 1 shows that in all instances, at least one SSLP was linked near a RGA. Also, SSLP090 and SSLP108 mapped in the vicinity of two framework markers linked to *Rsv4* (MLG D1b), a gene that confers resistance to all known strains of SMV (Hayes et al., 2000).

Putative Function of ESTs

Bioinformatics has allowed DNA sequence information to be related to protein information. Researchers can predict sequences that will code for particular proteins. Thus, a putative function can be assigned to a sequence based on the protein's function that it is predicted to generate. All of this information is accessible in the NCBI database.

A local database was created from soybean ESTs that contain such information. A BLAST search of the database with the 55 mapped ESTs produced homologous sequences for 36 of the ESTs. A putative function was assigned for 15 of these 36 ESTs (Table 4).

The protein sequences of the BLAST matches with our ESTs were analyzed for specific profiles or domains. Bioinformatics has allowed researchers to uncover patterns in protein sequences that will predict the formation of its coded protein. Combined studies have allowed structural domains to be categorized based on their involvement in biological processes. The domains and any informative descriptions about their functions are also reported in Table 2. Eleven of the 15 sequences scanned were related to defense responses in plants based on summaries provided by the programs used in the analysis. Ten of the ESTs will be discussed because they seemed to be directly related to disease resistance, all ten are highlighted red in Table 4. SSLP015 was the only sequence identified with a LRR domain, which is a common motif among disease resistance genes. The only marker identified as belonging to the chitinase family was SSLP039. SSLP161 contains an ethylene response domain that has been associated with a number of stress responses, such as fruit ripening, inhibition of stem and root elongation, and seed germination. SSLP021 was profiled to be involved with the removal of an initiator factor, methionine, from nascent proteins in the cytosol. SSLP059 has a domain known to protect cells from hydrogen peroxide. SSLP047 and SSLP061 have different domains but both were involved with strengthening of the cell wall. SSLP072 most likely plays a role in different methylation processes throughout the plant. SSLP082 is probably involved with various stress-induced regulations. It is a binding protein found throughout soybean organs that is also responsible for the regulation of temporal and developmental processes. Finally, SSLP147 belongs to a class of proteins that Westwood et al. (1998) have characterized with defense response to parasitization in plants. Understanding the processes the ESTs may govern offers two potential advantages. For one, it may lead to identifying which disease it provides resistance for based on the specifics of its activities. Secondly, it may provide insight to the interaction between the pathogen and gene (resistance pathway) if it is identified as a resistance locus.

Phylogenetic Analysis

Bioinformatics can be used to compare sequences and understand their origin of evolution. The sequences can be grouped within families to construct a phylogenetic tree. This type of analysis was performed on the 15 disease-related sequences that were assigned a putative function. The results are illustrated in Figure 2. Evolutionary history can provide insights into the development of resistance pathways. Related sequences may be involved in the same pathway or perform the same function in different pathways. If the function of one gene is known, it can possibly be used to identify new genes and sort out the involvement of other genes.

Related SSLPs can be determined from Figure 2, three main groupings were observed in the analysis. Each main group can be subcategorized into two groups that contains anywhere from 2 to 4 ESTs. Within these subcategories, the tree also allows the most related ESTs to be determined. The largest family consisted of SSLP048, SSLP047,

SSLP072, and SSLP061. SSLP082, SSLP015, and SSLP059 were also of close origin. It is possible that the markers included in each group have similar functions.

Discussion

As sequence information becomes more refined, molecular markers increase in their effectiveness. The number of ESTs being added to public databases rises exponentially every day. RGAs have been identified in a number of important crop species. These changes have led to new strategies that focus on targeting marker development for accelerated gene characterization and cloning. 47 ESTs showing similarity to disease resistance genes were mapped in *Arabidopsis* (Botella et al. 1997) and 109 ESTs presumed to be involved with some type of disease resistance were mapped in rice (Wang et al. 2001). Research in soybean has already mapped RGAs to locations in the genome associated with disease resistance that facilitated gene cloning (Hayes et al. 2000b; Gore et al. 2002; Hayes et al., in press). In the present study, 48 disease-related ESTs and 5 RGAs were integrated into a soybean linkage map. Several of the SSLPs mapped to regions known to contain disease resistance loci and QTLs. SSLPs were observed to map near all of the RGAs. This co-linearity between SSLPs and RGAs is evidence for functional resistance genes. One SSLP-RGA pair mapped to a region on chromosome I not yet identified as containing a resistance gene. The findings of this study should be useful for map-based cloning and promote the discovery of new resistance loci.

Six ESTs and four RGAs were directly linked to disease resistance genes, providing support for our markers to be employed in the candidate gene approach. For instance, on Figure 1, RGA43 and SSLP039 mapped to the bottom of chromosome J. This region possesses several resistance loci for the fungal pathogens, powdery mildew, phytopthora, and brown stem rot, and would be an excellent starting point for the candidate gene approach using our markers. The EST (SSLP039) belongs to the chitinase family, which is known to destroy the cell walls of invading plant pathogens. It is possible that the SSLP or RGA is part of one of the seven resistance genes at the bottom of MLF J. For RGA43, it may even be the entire gene.

The *Rsv1* gene on chromosome F that encodes resistance for SMV has been cloned at Virginia Tech (personal communication, Saghai Maroof). The *Rsv4* gene provides resistance to all strains of SMV, but has not yet been cloned. On chromosome D1b, SSLP090 and SSLP108 mapped to locations near this gene, providing additional markers for identification of the gene. The high-resolution mapping and chromosome walking methods employed to clone *Rsv1* can be facilitated with more markers like SSLP090 and SSLP108, if applied for cloning of *Rsv4*. Two reasons that relate directly to the candidate gene approach, separate our SSLPs from other markers; (1) the identification of markers that can be used for high-resolution mapping is much more efficient because they detect regions of the DNA that are contained within functional disease-related genes and (2) once a resistance locus becomes tightly flanked with markers, the sequence information that the SSLPs represent is already known.

SSLP015, which contains a domain (LRR) common to resistance genes, and RGA19a were associated with a region on chromosome E associated with a QTL providing resistance to SCN. QTLs are often difficult to track because so many genes are

involved. The addition of EST markers to the SCN QTL region on chromosome E, along with the SSLPs and RGAs shown on Figure 1 in this region, will provide tools which may be helpful in pinpointing the location of this gene and organizing the effects of QTLs conferring resistance to SCN. QTLs have been difficult to monitor because of their additive nature and the large role the environment can play on their abilities. We have demonstrated an efficient method for identifying markers related to QTLs, which is the only way to track and understand all of the genes participating in the defense response.

The highest concentration of SSLPs was on chromosome D1a and L, each linked to 6 markers. D1a has QTLs for SCN spread throughout the chromosome, which may be a reflection of the SSLPs linked and ordered on this group. However, chromosome L has no reported regions associated with disease resistance according to Soybase. Our observations support the possible uncovering of a new disease resistance region on linkage group L.

Understanding genome organization at the molecular level is a new trend in plant research that has been initiated by genomics and its techniques. The hope is to unravel resistance pathways and where their genetic information is stored. Characterization at this level will allow more controlled manipulation and classification of resistance responses. Categorizing resistance genes would ease the unveiling of related genes that have yet to be discovered. The nature of the molecular markers used in this study allowed the opportunity to try to monitor the organization of related genes that may perform the same function in different pathways. This task was attempted by defining the function of ESTs, relating them to each other through a phylogenetic analysis, and analyzing their distribution throughout the soybean genome.

Studying the evolution of genes allows them to be grouped according to their similarity. Genes mutate to acquire needed functions, so the closer in origin two genes are the fewer mutations (differences) will be present. Also, the functions of the two genes should become more related with a closer origin. A phylogenetic analysis can be performed to construct a "family tree" where the most divergent genes will be farthest from each other. SSLP015 and SSLP039 were the most evident ESTs to be resistance genes. SSLP015 contains an LRR domain that is known to be a motif in resistance genes. SSLP039 is related to production of chitinase which is also known to be directly involved with resistance responses by breaking down the cell walls of fungal pathogens. In our phylogenetic analysis, SSLP021 (translation initiation factor) was closely related to SSLP039 (acidic chitinase), while SSLP059 (catalase) and SSLP082 (glucosidase) were closely related to SSLP015 (LRR). Interestingly, SSLP021a and SSLP082 are located on linkage group L that was discussed above as being concentrated with disease-related ESTs and possible linkage with QTL resistance. Perhaps this region of chromosome L contains one or more QTLs for some type of fungal pathogen. Further phylogenetic analysis of the 15 profiled ESTs with the remaining ESTs (SSLPs) that markers were designed from may allow more SSLPs to be added to the related groupings identified in this study. The results could lead to the identification of more ESTs and their function.

The second part to gene evolution is figuring out how the genome has arranged sequences that store important genetic information. How have genes been packaged so they can survive, but still evolve to fulfill the survival of the fittest? ESTs that were identified with similar domains and functions were also most related in the phylogenetic

analysis. These results would be expected and are congruent for reliability of our EST identification. No related ESTs mapped near each other, and the location of all ESTs seemed to be randomly distributed if resistance loci are not taken into account. A possible explanation is preservation of the function. If all genes performing related functions were isolated, then the function would be lost if anything destroyed or changed that region in the genome. The organism may also want to keep all genes participating in a specific biological process together. SSLP082, SSLP161, and SSLP021A have different roles in a plant's defense response to disease and are all linked to chromosome L. It is possible that these three ESTs interact with each other to provide some type of disease resistance. Further knowledge of resistance pathways is required to make conclusions.

Resistance loci have successfully been isolated using the candidate gene approach and high resolution mapping. The results of our research have identified ESTs and/or RGAs that should be useful for the cloning of the *Rsv4* SMV resistance gene and a region on MLG J clustered with fungal resistance loci for Brown Stem Rot, Phytopthora, Nodulation, Powdery Mildew, and Frogeye Leaf Spot. However, there is no documentation of cloning a QTL for disease resistance in plants. Genetic analysis of the small and quantitative effects associated with QTLs is difficult. The type of markers used in this study is an excellent source for representing resistance genes involved with QTLs. Mapping many ESTs to a QTL and researching the functions of the sequences used would be especially beneficial to unraveling the mysteries of these regions. Our findings warrant further study to see if the effects of the QTLs can be more precisely monitored using our markers. Eventually, cloning of the QTLs should be possible. Finally, a couple of locations, particularly on linkage groups I and L, contained clusters of our SSLP markers, possibly indicating the presence of defense response genes. It would be interesting to see if the markers could be related to the presence of any disease resistance soybean genes on these chromosomes. Segregation analysis of our markers with different disease resistance genes may provide answers to some of these questions.

The SSLPs and RGAs used in this study provided a quick and easy way to identify and integrate disease related markers into a soybean linkage map. The markers can be used to define any where from 400 to 600 nucleotides of functional sequence information. Primer sequences can be shared among labs, so that our results can be transferable to other studies. Marker design of ESTs and RGAs is an excellent method for linkage map construction and gene identification.

References

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

Ashfield, T., Dazer, J., Held, D., Clayton, D., Keim, P., Maroof, M., Webb, D., and Innes, R. *Rpg1*, a soybean gene effective against races of bacterial blight, maps to a cluster of previously identified disease resistance genes. Theor. Appl. Genet. **96**: 1013-1021.

Bachman, M., Tamulonis, J., Nickell, C., and Bent, A. 2001. Molecular markers linked to brown stem rot resistance genes, *Rbs1* and *Rbs2*, in soybean. Crop Science **41:**527-535.

Bligh, H.F.J., Blackhall, N.W., Edwards, K.J., and McClung, A.M. 1999. Using amplified fragment length polymorphisms and simple sequence length polymorphisms to identify cultivars of brown and white milled rice. Crop Sci. **39**: 1715-1721.

Botella, M.A., Coleman, M.J., Hughes, D.E., Nishimura, M.T., Jones, J.D., and Somerville, S.C. 1997. Map positions of 47 *Arabidopsis* sequences with sequence similarity to disease resistance genes. Plant Journal **12**: 1197-1211.

Chen, X., Temnykh, S., Xu, Y., Cho, Y.G., McCouch S.R. 1997. Development of a microsatellite framework map providing genome-wide coverage in rice (*Oryza sativa* L.). Theor Appl. Genet. **95**: 553-567.

Cho, Y.G., McCouch, S.R., Kuiper, M., Kang, M.R., Pot, J., Groenen, J.T.M., and Eun, M.Y. 1998. Integrated map of AFLP, SSLP, and RFLP markers using a recombinant inbred population of rice (*Oryza sativa* L.). Theor. Appl. Genet. **97**: 370-380.

Cicek, M. 1997. "Genetic analysis of quantitative trait loci associated with seed sucrose content using molecular markers in an interspecific *Glycine* cross." M.S. Thesis, Department of Crop and Soil and Environmental Sciences, Virginia Tech, Blacksburg, VA.

Collins, N., Park, R., Spielmeyer, W., Ellis, J., and Pryor, A.J. 2001. Resistance gene analogs in barley and their relationship to rust resistance genes. Genome **44**: 375-381.

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

Donald, T.M., Pellerone, F., Adam-Blondon, A.-F., Bouquet, A., Thomas, M.R., and Dry, L.B. 2002. Identification of resistance gene analogs linked to a powdery mildew resistance locus in grapevine. Theor. Appl. Genet. **104**: 610-618.

Fasoula, V., Harris, D., Bailey, M., Phillips, D., and Boerma, H. 2003. Identification, mapping, and confirmation of a soybean gene for *bud blight* resistance. Crop Sci. **43**: 1754-1759.

Felsenstein, J. 1993. PHYLIP (Phylogeny Inference Package) version 3.5c. Distributed by the author. Department of Genetics, University of Washington, Seattle.

Fourmann, M., Charlot, F., Froger, N., Delourme, R., and Brunel, D. 2001. Expression, mapping, and genetic variability of *Brassica napus* disease resistance gene analogues. Genome **44**: 1083-1099.

Fred Hutchinson Cancer Research Institute. 2001. "BLOCKS version 01." Seattle, Washington: Fred Hutchinson Cancer Research Institute, retrieved October 1, 2003. http://blocks.fhcrc.org/blocks/blocks_search.html.

Gore, M.A., Hayes, A.J., Jeong, S.C., Yue Y.G., Buss, G.R. and Saghia Maroof, M.A. 2002. Mapping tightly linked genes controlling potyvirus infection at the *Rsv1* and *Rpv1* region in soybean. Genome. **45**: 592-599.

Graham, M.A., Marek, L.F., Lohnes, D., Cregan, P., and Shoemaker, R.C. 2000. Expression and genome organization of resistance gene analogs in soybean. Genome **43**: 86-93.

Graham, M., Marek, L., and Shoemaker, R. 2002. PCR sampling of disease resistance like sequences from a disease resistance gene cluster in soybean. Theor. Appl. Genet. **105**: 50-57.

Hayes, A.J., Yue, Y.G., and Saghai Maroof, M.A. 2000a. Expression of two soybean resistance gene candidates shows divergence of paralogous single-copy genes. Theor. Appl. Genet. **101**: 789-795.

Hayes, A., Ma, G., Buss, G. and Maroof, M.A. 2000b. Molecular marker mapping of *Rsv4*, a gene conferring resistance to all known strains of soybean mosaic virus. Crop Science **40**: 1434-1437.

Holton, T.A., Christopher, J.T., McClure, L., Harker, N., and Henry, R.J. 2002. Identification and mapping of polymorphic SSR markers from expressed gene sequences of barley and wheat. Molecular Breeding **9**: 63-71.

Jeong, S., Kristipati, S., Hayes, A., Maughan, P., Noffsinger, S., Gunduz, I., Buss, G., and Saghai Maroof, M. 2002. Genetic and sequence analysis of markers tightly linked to the *soybean mosaic virus* resistance gene, *Rsv3*. Crop Sci. **42**: 265-274.

Kanazin, V., Marek, L.F., and Shoemaker, R.C. 1996. Resistance gene analogs are conserved and clustered in soybean. PNAS **93**: 11746-11750.

Kim, H., and Diers, B. 2000. Inheritance of partial resistance to sclerotinia stem rot in soybean. Crop Sci. **40**: 55-61.

Lohnes, D., and Schmitthenner, A. 1997. Position of the phytopthora resistance gene *Rps7* on the soybean molecular map. Crop Sci. **37**: 555-556.

Mago, R., Nair, S., and Mohan, M. 1999. Resistance gene analogues from rice: cloning, sequencing and mapping. Theor. Appl. Genet. **99**: 50-57.

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

Meksem, K., Pantazopoulos, P., Njiti, V., Hyten, L., Arelli, P., and Lightfoot, D. 2001. Forrest resistance to the soybean cyst nematode is bigenic: saturation mapping of *Rhg1* and *Rhg4* loci. Theor. Appl. Genet. **103**: 710-717.

Moncada, P., Martinez, C.P., Borrero, J., Chatel, M., Gauch, H., Guimaraes, E., Tohme, J., McCouch, S.R. 2001. Quantitative trait loci for yield and yield components in an *Oryza sativa x Oryza* BC_2F_2 population evaluated in an upland environment. Theor. Appl. Genet. **102**: 41-52.

Mudge, J., Cregan, P., Kenworthy, J., Kenworthy, W., Orf, J., and Young, N. 1997. Two microsatellite markers that flank the major soybean cyst nematode resistance locus. Crop Sci. **37**: 1611-1615.

Narvel, J., Jakkula, L., Phillips, D., Wang, T., Lee, S., and Boerma, H. 2001. Molecular mapping of *Rxp* conditioning reaction to bacterial pustule in soybean. J. Hered. **92**: 267-270.

NCBI. 2003. "BLASTX program." Maryland: National Center for Biotechnology Information, retrieved October 1, 2003. http://www.ncbi.nlm.nih.gov/BLAST/.

North Carolina Supercomputing Center. 2002. "Prosite version 17.2." North Carolina: North Carolina Supercomputing Center, retrieved October 1, 2003. http://us.expasy.org/prosite/.

Palmer, R., and Kilen, T. 1987. Qualitative genetics and cytogenetics. Agronomy **16**: 135-209.

Penuela, S., Danesh, D., and Young, N.D. 2002. Targeted isolation, sequence analysis, and physical mapping of nonTIR NBS-LRR genes in soybean. Theor. Appl. Genet. **104**: 261-272.

Polzin, K., Lohnes, D., Nickell, C., and Shoemaker, R. 1994. Integration of *Rps2, Rmd,* and *Rj2* into linkage group J of soybean molecular map. J. Hered. **85**: 300-303.

Prabhu, R., Njiti, V., Bell-Johnson, B., Johnson, J., Schmidt, M., Klein, J., and Lightfoot, D. 1999. Selecting soybean cultivars for dual resistance to soybean cyst nematode and sudden death syndrome using two DNA markers. Crop Science **39**: 982-987.

Rector, B., All, J., Parrott, W., and Boerma, H. 1999. Quantitative trait loci for antixenosis resistance to corn earworm in soybean. Crop Sci. **39**: 531-538.

Rongwen, J., Akkaya, M.S., Bhagwat, A.A., Lavi, U., and Cregan, P.B. 1995. The use of microsatellite DNA markers for soybean genotype identification. Theor. Appl. Genet. **90**: 43-48.

Saghai Maroof, M.A., Soliman, K.M., Jorgenson, R.A., and Allard, R.W. 1984. Ribosomal DNA spacer length polymorphism in barley: Mendelian inheritance, chromosomal location, and population dynamics. PNAS **81**: 8014-8018.

Scott, K.D., Eggler, P., Seaton, G., Rossetto, M., Ablett, E.M., Lee, L.S., and Henry, R.J. 2000. Analysis of SSRs derived from grape ESTs. Theor. Appl. Genet. **100**: 723-726.

San Diego Supercomputing Center. 2003. "Biology workbench." San Diego, California: University of California, San Diego, retrieved October 1, 2003. http://workbench.sdsc.edu/.

Shoemaker, R.C., Polzin, K.M., Lorenzen, L.L., and Specht, J.E. 1996. Soybean: Genetics, Molecular Biology, and Biotechnology: Molecular genetic mapping of soybean. Ed. Verma D.S., and Shoemaker, R.C. CAB INTERNATIONAL. UK.

Shoemaker et al. 2003. "Soybase class browser: Pathology." Iowa: Iowa State University, retrieved September 30, 2003. http://soybase.ncgr.org/cgi-bin/ace/generic/search/soybase.

Sinclair, J.B., and Shurtleff, M.C. 1999. Compendium of Soybean Diseases. (4^{th}ed.). The American Phytopathological Society, Inc. St. Paul, Minnesota.

Tamulonis, J., Luzzi, B., Hussey, R., Parrott, W., and Boerma, H. 1997a. DNA markers associated with resistance to javanese root-knot nematode in soybean. Crop Sci. **37**: 783-788.

Tamulonis, J., Luzzi, B., Hussey, R., Parrott, W., and Boerma, H. 1997b. DNA marker analysis of loci conferring resistance to peanut root-knot nematode in soybean. Theor. Appl. Genet. **95**: 664-670.

Tamulonis, J., Luzzi, B., Hussey, R., Parrott, W., and Boerma, H. 1997c. RFLP mapping of resistance to sourthern root-knot nematode in soybean. Crop Sci. **37**: 1903-1909.

Thibaud-Nissen, F., Shealy, R., Khanna, A., and Vodkin, L. 2003. Clustering of microarray data reveals transcript patterns associated with somatic embryogenesis in soybean. Plant Physiology. **132**: 118-136.

Thompson J.D., Higgins D.G., and Gibson T.J. 1994. "CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice." Nucleic Acids Res. 22:4673-4680.

University of Minnesota. 2003. "Metafam." Minnesota: University of Minnesota, retrieved October 1, 2003. http://metafam.ahc.umn.edu/.

Wang, Z., Taramino, G., Yang, D., Liu, G., Tingey, S.V., Miao, G.H., and Wang, G.L. 2001. Rice ESTs with disease-resistance gene or defense-response gene-like sequences mapped to regions containing major resistance genes or QTLs. Mol. Genet. Genomics **265**: 302-310.

Washington University. 2003. "Prints version 32." St. Louis, Missouri: Washington University, retrieved October 1, 2003. http://bioinf.man.ac.uk/cgi-bin/dbbrowser/fingerPRINTScan/muppet/FPScan.cgi.

Westwood, J.H., X. Yu, C.L. Foy and C.L. Cramer. 1998. Expression of a defense-related 3-hydroxy-3-methylglutaryl CoA reductase gene in response to parasitization by Orobanche. Mol. Plant-Microbe Interact 11:530-536.

Yang, W., Weaver, D., Nielson, B., and Qui, J. 2001. Molecular mapping of a new gene for resistance to frogeye leaf spot of soybean in Peking. Plant Breeding **120**: 73-78.

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

Yue, P., and Sleper, D. 2001a. Molecular characterization of resistance to *heterodera glycines* in soybean PI 438489B. Theor. Appl. Genet. **102**: 921-928.

Yue, P., Sleper, D., and Arelli, P. 2001b. Mapping resistance to multiple races of *heterodera glycines* in soybean PI 89772. Crop Sci. **41**: 1589-1595.

Table 1. Markers that could not be amplified without modifying standard PCR conditions.

Marker	Annealing temp. (°C)[1]	# of Cycles[2]
SSLP008	43	40
SSLP009	43	40
SSLP016	47[+]	40
SSLP021a[3]	47	42
SSLP021b[3]	47	42
SSLP072	47	42
SSLP082	47	42
SSLP115	47	40
SSLP181	47	40
Sat069	47	42
Sat104	47	40
Satt170	47	40
Satt281	47	40
Satt313	47	40

[1]The temperature that is set for the step that includes annealing of primers to the DNA and amplification of it. Decreasing annealing temperature will reduce the specificity of the primers and allow them to bind DNA easier. Thus, amplification products can be obtained from primers that would otherwise not bind to DNA.

[2]Indicates how many times the amplification reaction was allowed to repeat itself. The number of cycles can be increased to allow primers that don't amplify well to produce more amplification product.

[3]Lowercase letters indicate that the marker amplified two different regions of the genome. The letters were used to differentiate the regions.

[+]This annealing temperature was used for standard PCR conditions (47°C). The number of cycles was changed for these markers.

Table 2. Known Disease Resistance Genes Superimposed on Our Skeletal Frame Map of Published Markers.

Disease Resistance[1]	Gene	MLG	Reported Markers[2]	Reference
SCN (minor QTL)	-	A2	Binned with Satt187	Shoemaker et al., 2003*
SMV	*Rsv3*	B2	A519 (0.9cM)	Jeong et al., 2002
CE (minor QTL)	-	C2	B131_1	Rector et al., 1999
SCN (4 minor QTLs)	-	D1a	Binned with Satt184 and Satt368	Yue et al., 2001a
SMV	*Rsv4*	D1b	Satt542 (4.7cM) Satt558 (7.8cM)	Hayes et al., 2000
Bacterial Pestule	*Rxp*	D2	Satt372 (3.9cM) Satt014 (12.4cM)	Narvel et al., 2001
SCN QTL 18-4^{+}	-	E	Binned with Satt452	Shoemaker et al., 2003
PRN (minor QTL)	-	E	Binned with A111H_2 and Satt231	Tamulonis et al., 1997b; Shoemaker et al., 2003
JRN QTL 1-2^{+}	-	F	Binned with RFLP A806_1	Shoemaker et al., 2003
Peanut Mottle Virus	*Rpv1*	F	Hsp176, Satt510, Satt120, and *Rsv1*	Gore et al., 2002
SMV	*Rsv1*	F	*Rpv1*, Hsp176, Satt510, and Satt120	Gore et al., 2002
Bacterial Blight	*Rpg1*	F	php2265 and php2385	Ashfield et al., 1998
Bud Blight (major QTL)	-	F	Satt510	Fasoula et al., 2003
CE (minor QTL)	-	F	B212V_1	Rector et al., 1999
JRN (major QTL)	-	F	B212V_1	Tamulonis et al., 1997a
PRN (major QTL)	-	F	B212V_1	Tamulonis et al., 1997b
SCN (major QTL)	-	G	Satt038 (3cM) Satt309 (2cM)	Mudge et al., 1997; Yue et al., 2001b
SCN (minor QTL)	-	G	Satt309 (4cM)	Yue et al., 2001b
Sudden Death Syndrome	*Rfs*	G	Satt038	Prabhu et al., 1999
SKN (minor QTL)	-	G	K493_1	Tamulonis et al., 1997c
CE (minor QTL)	-	H	R249_2	Rector et al., 1999
Frogeye Leaf Spot	*Rcs3*	J	Satt244 (1.1cM)	Yang et al., 2001
Brown Stem Rot	*Rbs1*	J	Satt215 and Satt431	Bachman et al., 2001
Brown Stem Rot	*Rbs2*	J	Satt244 and Satt431	Bachman et al., 2001
Brown Stem Rot	*Rbs3*	J	Binned with *Rbs1* and *Rbs2*	Shoemaker et al., 2003
Powdery Mildew	*Rmd*	J	Satt431	Polzin et al., 1994; Graham et al., 2002
Phytopthora	*Rps2*	J	*Rmd*	Polzin et al., 1994; Graham et al., 2002
Nodulation	*Rj2*	J	*Rmd*	Polzin et al., 1994; Graham et al., 2002

Table 2 continued, Known Disease Resistance Genes Superimposed on Our Skeletal Frame Map of Published Markers.

Disease Resistance[1]	Gene	MLG	Reported Markers[2]	Reference
SR (minor QTL)	-	K	Satt046	Kim et al., 2000
SCN QTL 3-3[+]	-	M	Binned with Satt590	Shoemaker et al., 2003
CE (minor QTL)	-	M	A584_3	Rector et al., 1999
Phytopthora	*Rps7*	N	R022_1 (1.1cM)	Lohnes et al., 1997
Bacterial Blight	*Rpg4*	N	Satt521 (15.1cM)	Palmer et al., 1987; Shoemaker et al., 2003
SKN (major QTL)	-	O	G248_1	Tamulonis et al., 1997c

Abbreviations: (SCN) soybean cyst nematode; (SMV) soybean mosaic virus; (JRN) javanese root-knot nematode; (PRN) peanut root-knot nematode; (SR) *sclerotinia* stem rot; (SKN) southern root-knot nematode; (CE) corn earworm; (MLG) molecular linkage group.

[1]The type of disease resistance that is provided by the gene. For QTLs, their effect is additive, so a single gene was not reported. If the literature indicates that the QTL explains 30% or more of phenotypic variation, then it was labeled as major. QTLs explaining less than 30% of phenotypic variation were labeled as minor.

[+]QTL numbers correspond to the identification used by Soybase (Shoemaker et al., 2003).

[2]The disease resistance genes were superimposed on our linkage map by associating reported markers linked to them with the published markers used to construct the skeletal frame of our map. Specific linkage distances that were reported between the markers and the resistance genes are recorded in parenthesis and are in centimorgans (cM).

[*]Shoemaker et al. (2003) have created a website called Soybase that contains information about disease resistance loci for each MLG. Bins are created that represent regions of a MLG, so that resistance genes and molecular markers from the same region can be grouped together. However, the precise location of the genes is not defined very well.

Table 3. The 48 SSLPs and 5 RGAs that were integrated into our skeletal frame map that was associated with known disease resistance genes.

Marker ID	MLG	Associated Resistance gene	Framework		Reported Marker	
			Marker	Distance (cM)	Marker	Distance (cM)
SSLP138	A2	-	Satt158	8.4	-	-
SSLP060	A2	-	Satt455	3.0	-	-
SSLP124	B1	-	Satt509	15.3	-	-
SSLP061	B1	-	Sat123	7.7	-	-
SSLP114	B1	-	Sat123	2.5	-	-
SSLP147	B2	-	Satt467	25.3	-	-
SSLP143	C1	-	Satt578	0.6	-	-
Sat085^{+}	C1	-	NA	NA	-	-
SSLP094	C1	-	Satt476	11.7	-	-
SSLP116	C1	-	Satt180	0.0	-	-
Sat062^{+}	C2	-	NA	NA	-	-
SSLP009	C2	-	Satt281	0.2	-	-
SSLP024	C2	-	Satt170	7.8	-	-
SSLP102	C2	-	Satt316	15.7	-	-
SSLP057	D1a	-	Satt531	24.4	-	-
SSLP022	D1a	*SCN QTL*-	Satt368	2.0	Satt368	2.0 / Binned
Satt368^{+}	D1a	*SCN QTL*	NA	NA	Satt368	0.0 / Binned
SSLP117	D1a	-	Sat036	2.0	-	-
SSLP130	D1a	-	Sat036	7.1	-	-
Satt407^{+}	D1a	-	NA	NA	-	-
SSLP048	D1b	-	Satt216	2.8	-	-
SSLP090	D1b	*Rsv4*	Satt558	0.0	Satt558	0.0 / 7.8
SSLP108	D1b	*Rsv4*	Satt542	8.3	Satt542	8.3 / 4.7
SSLP016	D2	-	Sat022	6.1	-	-
RGA19a	E	*SCN QTL*	Satt598	0.0	Satt452	10.0 / Binned
SSLP015	E	*SCN QTL*	Satt598	2.3	Satt452	10.0 / Binned
SSLP163	E	-	Satt045	35.4	-	-
SSLP079	F	*JRN QTL*	Sat133	3.9	A806_1	14.0 / Binned
SSLP091	F	*JRN QTL*	Sat133	3.9	A806_1	14.0 / Binned
SSLP142	G	-	Satt235	26.2	-	-
SSLP017	G	*Rfs and SCN QTL*	Satt235	2.8	Satt038	15.0 / 3.0
SSLP193	H	-	Satt052	9.9	-	-
SSLP115	I	-	Satt367	7.3	-	-
SSLP018	I	-	Satt239	0.0	-	-
RGA18	I	-	Satt239	4.4	-	-
Sat105^{+}	I	-	NA	NA	-	-
SSLP013	I	-	Sct189	3.8	-	-
SSLP181	J	-	Satt529	17.7	-	-

Table 3 continued, The 48 SSLPs and 5 RGAs that were integrated into our skeletal frame map that was associated with known disease resistance genes.

Marker ID	MLG	Associated Resistance gene	Framework		Reported Marker	
			Marker	Distance (cM)	Marker	Distance (cM)
SSLP039	J	*Rcs3, Rbs1, Rbs2, and Rbs3*	Satt244	0.0	Satt244	0.0 / 1.1
RGA43	J	*Rmd, Rps2, and Rj2*	Satt431	3.2	Satt431	0.0 / Linked
SSLP008	K	*SR QTL*	Satt046	22.3	Satt046	0.0 / Linked
SSLP021a	L	-	Satt495	6.0	-	-
SSLP083	L	-	Satt495	6.1	-	-
SSLP023	L	-	Satt495	10.0	-	-
Satt398[+]	L	-	NA	NA	-	-
SSLP082	L	-	Satt398	11.9	-	-
SSLP161	L	-	Sat113	0.7	-	-
SSLP047	M	*SCN QTL*	Satt150	7.7	Satt590	9.0 / Binned
RGA19b	M	*SCN QTL*	Satt150	16.5	Satt590	9.0 / Binned
RGA307	M	*SCN QTL*	Satt150	16.5	Satt590	9.0 / Binned
SSLP053	M	*CE QTL*	Satt536	3.7	A584_3	4.0 / Linked
SSLP100	O	-	Satt478	15.9	-	-
SSLP002	O	-	Satt478	2.6	-	-

Abbreviations: (SCN) soybean cyst nematode; (SMV) soybean mosaic virus; (JRN) javanese root-knot nematode; (PRN) peanut root-knot nematode; (SR) *sclerotinia* stem rot; (SKN) southern root-knot nematode; (CE) corn earworm; (MLG) molecular linkage group; (NA) not applicable.

Associated Resistance Genes = Resistance loci that were superimposed on our skeletal map by their association with framework markers (Table 2), have now been associated with our markers that are linked to framework markers.
(-) = Although the SSLP was integrated into the skeletal map, it could not be closely related to any resistance gene. The column does not apply to the marker.
Framework = This column gives the framework marker that was linked to our marker, allowing it to be integrated into the skeletal map. Linkage distances are reported between the published (framework) marker and our marker.
Reported Marker = This column explains how our markers are associated with the disease resistance genes superimposed on the skeletal map (Table 2). The first number given in the distance column is an estimated distance between the reported marker, related to the resistance loci, and the framework marker. The second number given is an estimated distance between the reported marker and the resistance gene. (Binned) is taken from Table 2, and is a grouping used by Soybase that does not give precise locations. (Linked) indicates that the marker has been related to a region near the resistance loci, but exact distances were not reported.

+ Sequence information from these published markers corresponded to our disease-related ESTs. Thus, these markers were used for skeletal framework and to represent SSLPs from our ESTs. The two columns for "Marker Distance[1]" are not applicable for these markers.

Table 4. Description of 15 mapped ESTs (SSLPs) that were characterized using Bioinformatics tools to determine their putative function.

EST	Chr.	Putative Funct.	Domain/Profile[1]	Description[2]
SSLP013	I	Calmodulin	EF hand family	Binds calcium and was similar to calfagen that recognizes protozoa parasites in animals
SSLP015	E	Protein	LRR	Common motif in resistance genes
SSLP021	L	Translation initiation factor	Methionine aminopeptidase	Blocks action of initiator factor, methionine
SSLP039	J	Acidic chitinase	Chitinase family	Chitinase breaks down cell walls of pathogens
SSLP047	M	Proline rich protein	Proline rich extension signature	Strengthens cell walls
SSLP048	D1b	Arabinogalactan-protein precursor	None	None
SSLP059	-	Catalase	Catalase proximal heme-ligand	Guards against hydrogen peroxide
SSLP061	B1	Non-specific lipid transfer protein	Plant lipid transfer	Wax deposition, builds up cell walls
SSLP072	-	Adenosylhomocysteinase	S-adenosyl-L-homocysteine	Functions in methylation
SSLP082	L	Glucan endo-1,3-beta-glucosidase	Glycosyl hydrolases	Binding protein, regulates various stress-induced pathways
SSLP091	F	Peroxidase ATP26a	Peroxidase proximal heme-ligand	Plant peroxidase
SSLP142	G	Related protein isolog	None	None
SSLP143	C1	Serine/threonine protein kinase	Serine/threonine protein kinase	Regulates response to external stimuli
SSLP147	B2	Hydroxy-methylglutaryl-coenzyme A reductase	Hydroxy-methylglutaryl-coenzyme A reductase	Similar to proteins that control cholesterol
SSLP161	L	Ethylene response factor	Ethylene responsive element	Associated with various stress responses

(Chr.) =The chromosome the marker was mapped to is indicated in this column.
(-) =indicates that the chromosomal location is not known.
SSLPs highlighted in red were determined to be disease related.

[1]Sequences were scanned using several Bioinformatics tools (as reported in the Materials and Methods section: *Identifying EST function.*) that can classify them into families. The families are based on shared sequence information (motifs, predicted protein domains, etc.) that indicate specific functions.

[2]Many different programs were used in the analysis that provided their own descriptions for the function of a sequence and its involvement in biological pathways. Input from all programs was taken into account and used to develop the descriptions we have provided about the function of each EST.

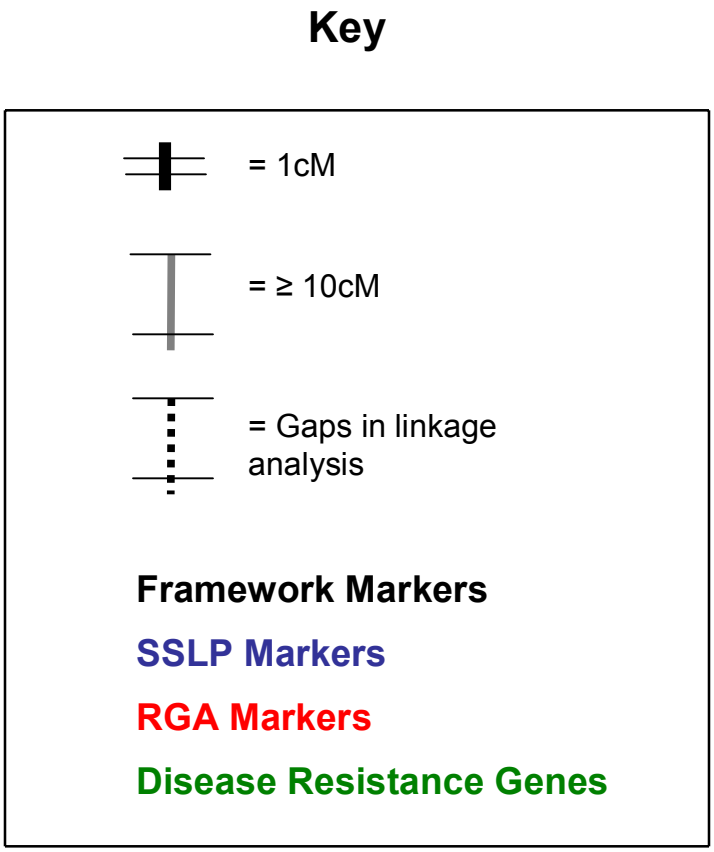


Figure 1. Genetic linkage map for soybean based on published framework (Cregan et al., 1999), SSLP, and RGA markers mapped in an F8 V71-370 X PI407162 RIL population. Twenty linkage groups are reported (**A1 – O**), map distances are given in centimorgans (cM) to the left of each linkage group.

[a] Represents the scale that was used to report linkage distance between markers. Solid black lines indicate that map scale was used

[b] Some linkage distances were too large to use map scale. To allow these distances to fit on our map, they were rescaled to the size indicated in the key and represented with grey lines.

[c] Dashed lines indicate that our linkage analysis could not group the markers together. Distances for the gaps have been estimated.

[d] Color code was used to identify what type of markers is reported. Published **Framework Markers** are highlighted black and have been used extensively in several different mapping populations and disease resistance studies. **SSLP Markers** are highlighted blue and **RGA Markers** are highlighted red. Published **Disease Resistance Genes** are superimposed on the map to the right of each linkage group and are highlighted in green.

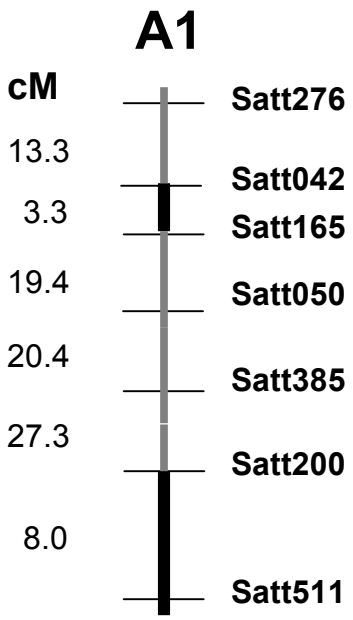


Figure 1 continued Genetic linkage map for soybean based on published framework (Cregan et al., 1999), SSLP, and RGA markers.

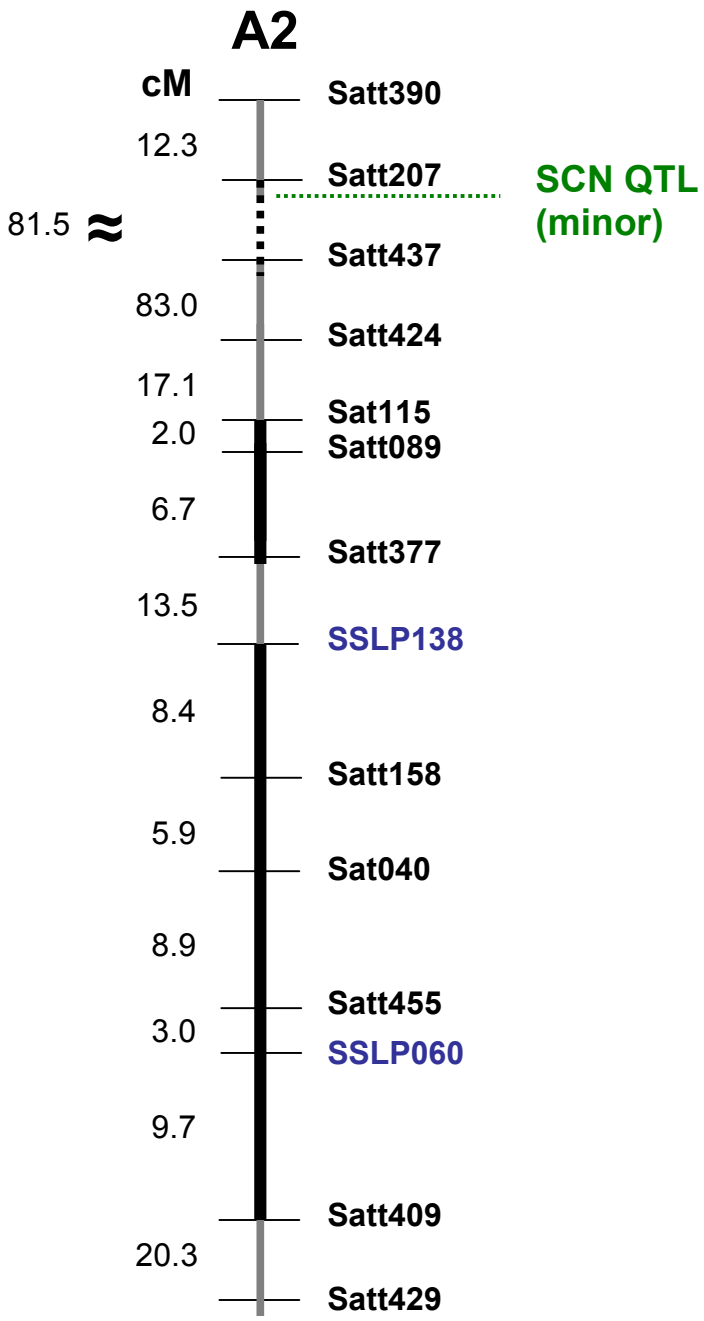


Figure 1 continued Genetic linkage map for soybean based on published framework (Cregan et al., 1999), SSLP, and RGA markers.

≈ Estimation for the gap distance is given to the left of the symbol. Estimations are based on marker location in several different mapping populations (Cregan et al., 1999).

QTL If literature indicates the QTL is responsible for 30% or more of phenotypic variation then it is considered major, less than 30% is considered minor.

Abbreviations: (SCN) soybean cyst nematode

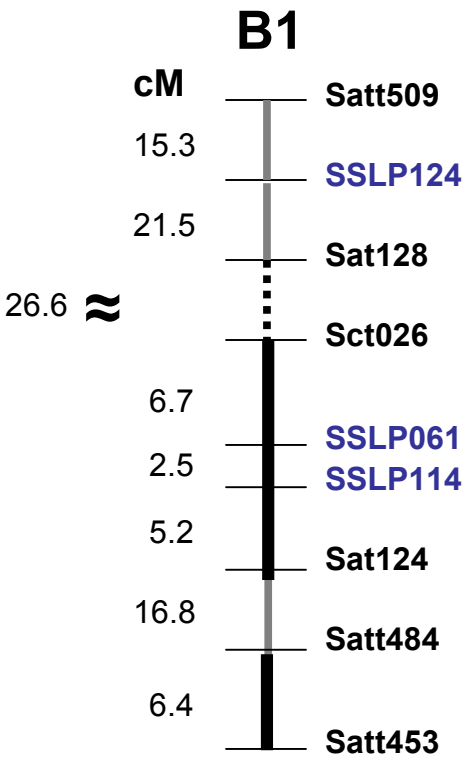


Figure 1 continued Genetic linkage map for soybean based on published framework (Cregan et al., 1999), SSLP, and RGA markers.

≈ Estimation for the gap distance is given to the left of the symbol. Estimations are based on marker location in several different mapping populations (Cregan et al., 1999).

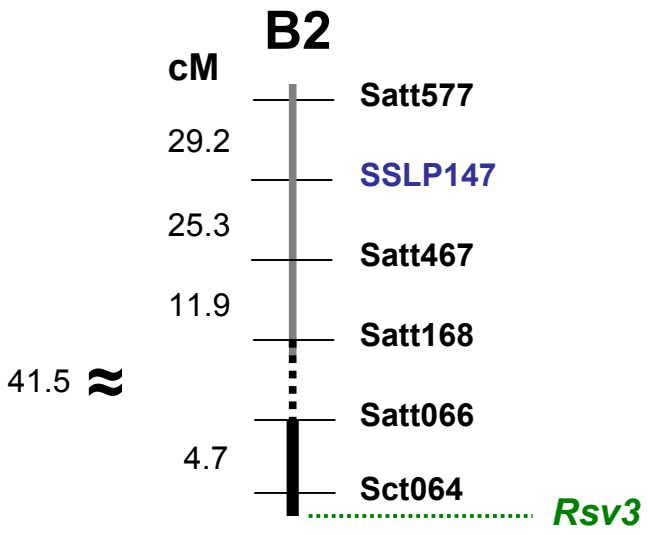


Figure 1 continued Genetic linkage map for soybean based on published framework (Cregan et al., 1999), SSLP, and RGA markers.

≈ Estimation for the gap distance is given to the left of the symbol. Estimations are based on markers location in several different mapping populations (Cregan et al., 1999).

Rsv3 = Soybean mosaic virus resistance gene

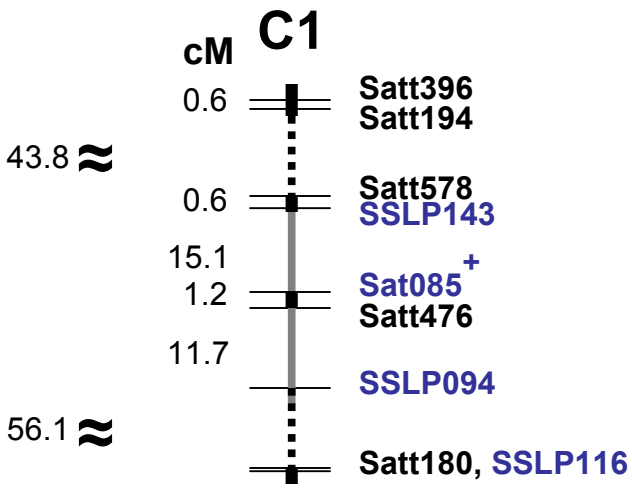


Figure 1 continued Genetic linkage map for soybean based on published framework (Cregan et al., 1999), SSLP, and RGA markers.

≈ Estimation for the gap distance is given to the left of the symbol. Estimations are based on markers location in several different mapping populations (Cregan et al., 1999).

+ The marker has dual roles. It is a published framework marker; however, it is highlighted blue because its sequence information corresponds to a polymorphic site that was identified in one of our disease-related ESTs. Thus, it also represents an additional SSLP.

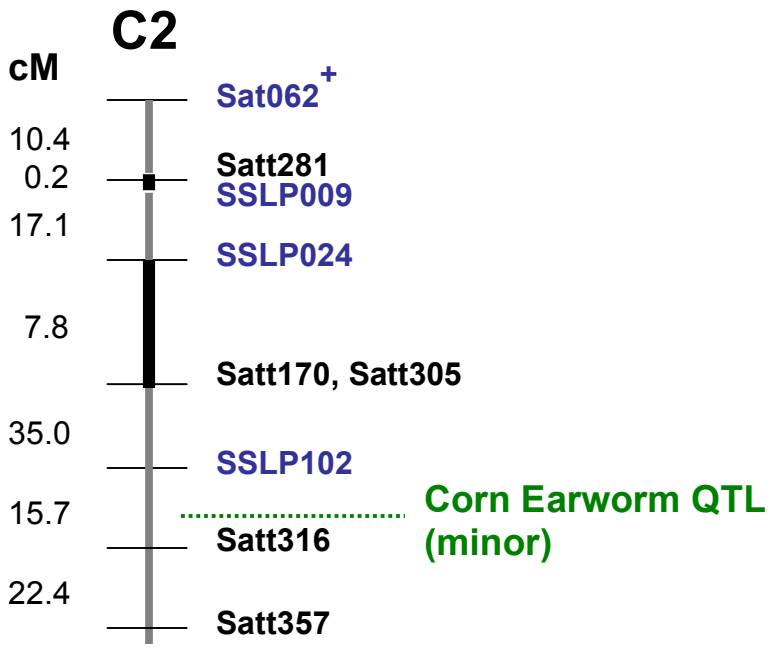


Figure 1 continued Genetic linkage map for soybean based on published framework (Cregan et al., 1999), SSLP, and RGA markers.

+ The marker has dual roles. It is a published framework marker; however, it is highlighted blue because its sequence information corresponds to a polymorphic site that was identified in one of our disease-related ESTs. Thus, it also represents an additional SSLP.

QTL If literature indicates the QTL is responsible for 30% or more of phenotypic variation then it is considered major, less than 30% is considered minor.

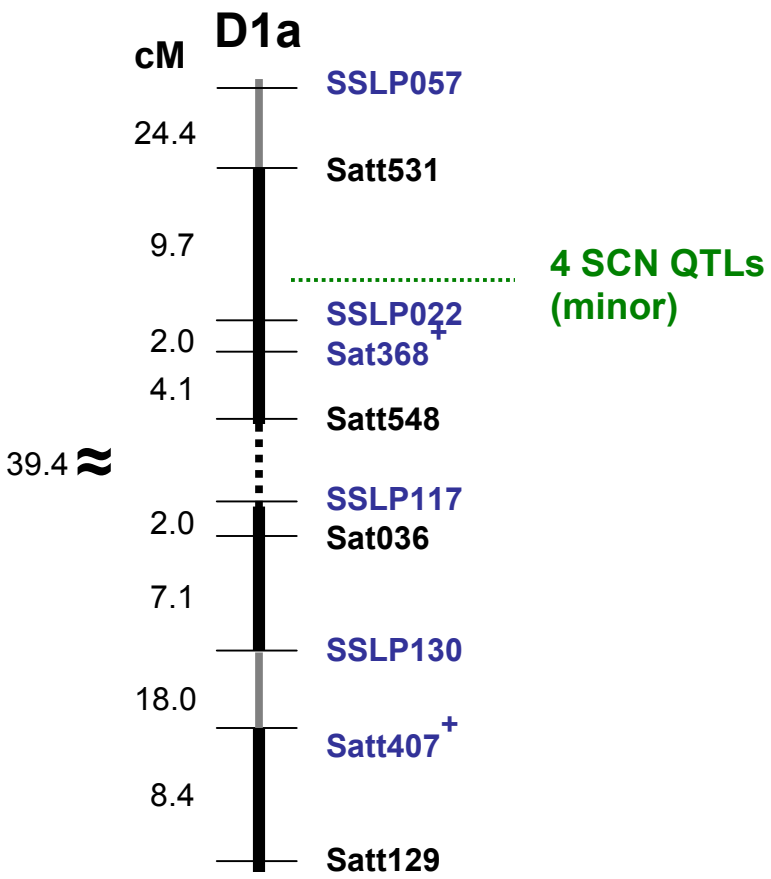


Figure 1 continued Genetic linkage map for soybean based on published framework (Cregan et al., 1999), SSLP, and RGA markers.

≈ Estimation for the gap distance is given to the left of the symbol. Estimations are based on markers location in several different mapping populations (Cregan et al., 1999).

+ The marker has dual roles. It is a published framework marker; however, it is highlighted blue because its sequence information corresponds to a polymorphic site that was identified in one of our disease-related ESTs. Thus, it also represents an additional SSLP.

QTL If literature indicates the QTL is responsible for 30% or more of phenotypic variation then it is considered major, less than 30% is considered minor.

Abbreviations: (SCN) soybean cyst nematode

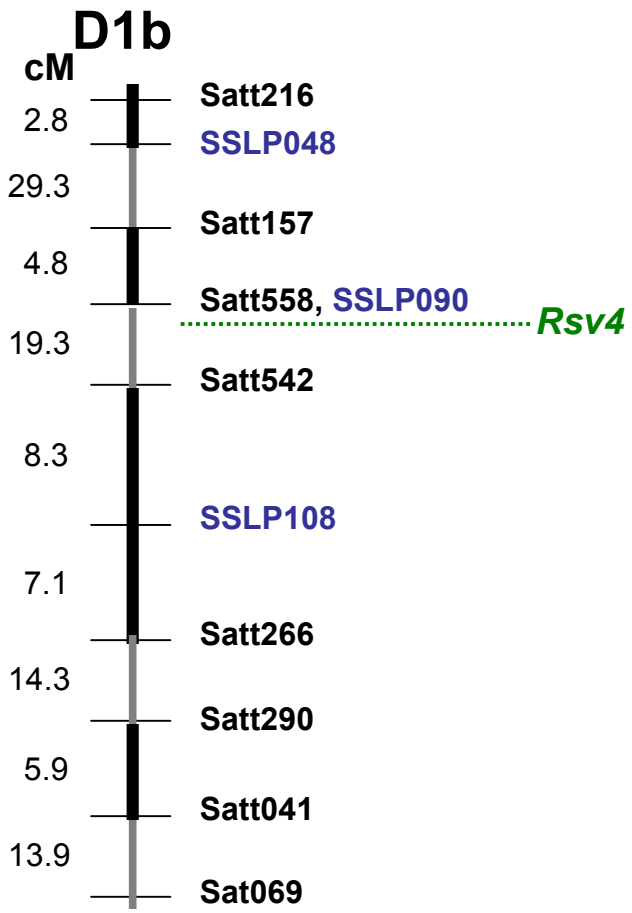


Figure 1 continued Genetic linkage map for soybean based on published framework (Cregan et al., 1999), SSLP, and RGA markers.

Rsv4 = Soybean mosaic virus resistance gene

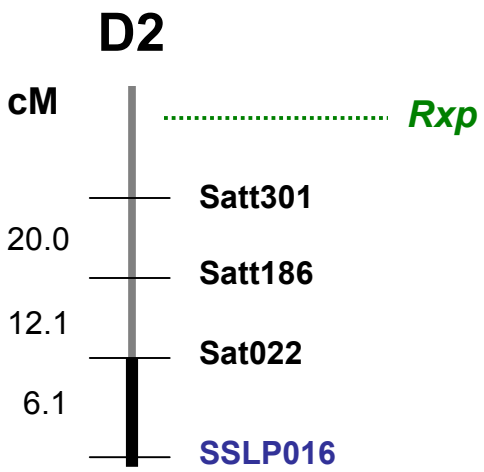


Figure 1 continued Genetic linkage map for soybean based on published framework (Cregan et al., 1999), SSLP, and RGA markers.

Rxp = Bacterial Pestule resistance gene

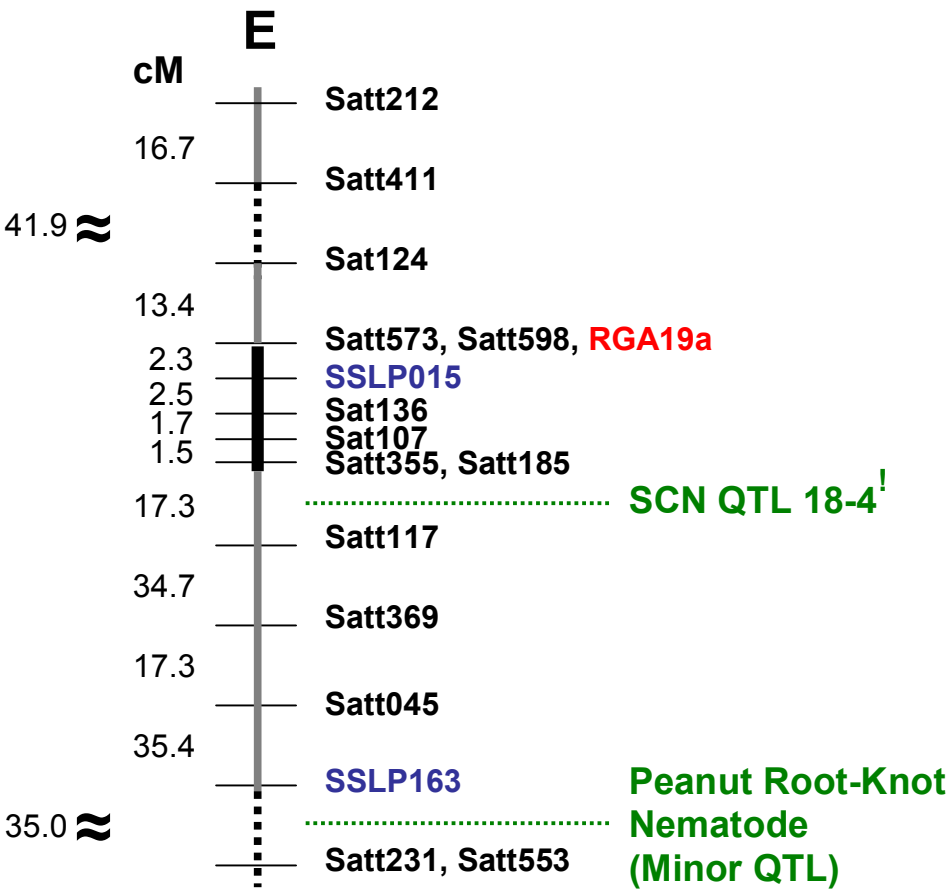


Figure 1 continued Genetic linkage map for soybean based on published framework (Cregan et al., 1999), SSLP, and RGA markers.

≈ Estimation for the gap distance is given to the left of the symbol. Estimations are based on markers location in several different mapping populations (Cregan et al., 1999).

! QTL name as reported by Soybase (Shoemaker et al., 2003).

QTL If literature indicates the QTL is responsible for 30% or more of phenotypic variation then it is considered major, less than 30% is considered minor.

Abbreviations: (SCN) soybean cyst nematode

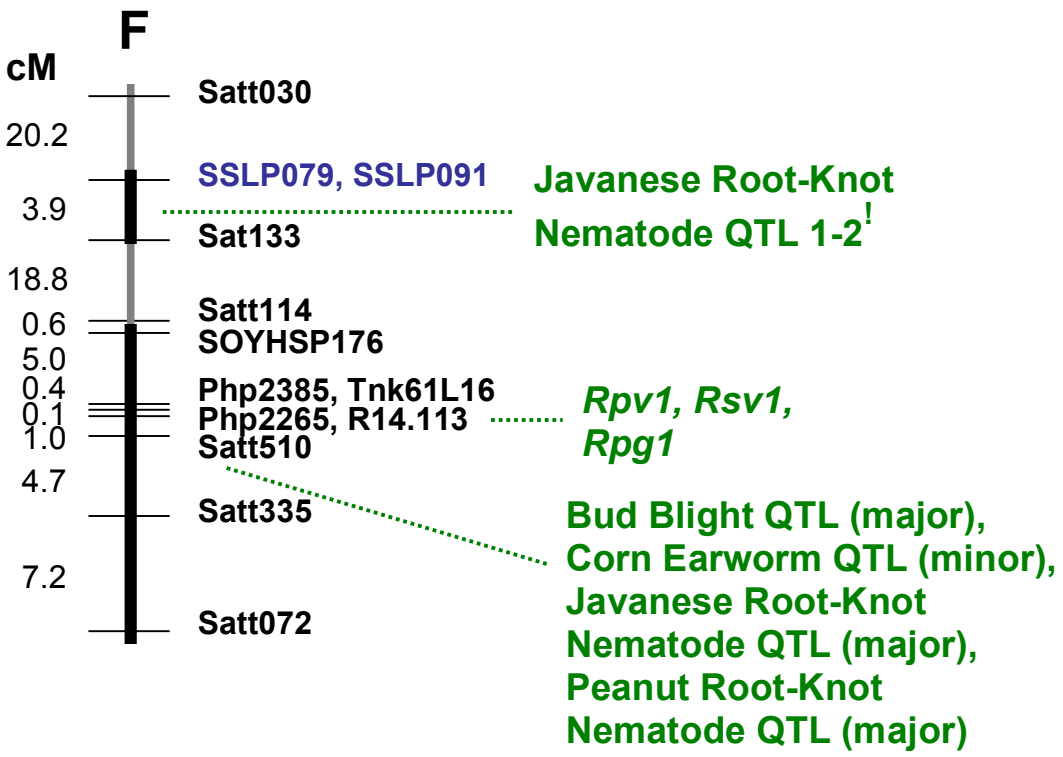


Figure 1 continued Genetic linkage map for soybean based on published framework (Cregan et al., 1999), SSLP, and RGA markers.

! QTL name as reported by Soybase (Shoemaker et al., 2003).

QTL If literature indicates the QTL is responsible for 30% or more of phenotypic variation then it is considered major, less than 30% is considered minor.

Rsv1 = Soybean mosaic virus resistance gene
Rpv1 = Peanut mottle virus resistance gene
Rpg1 = Bacterial Blight resistance gene

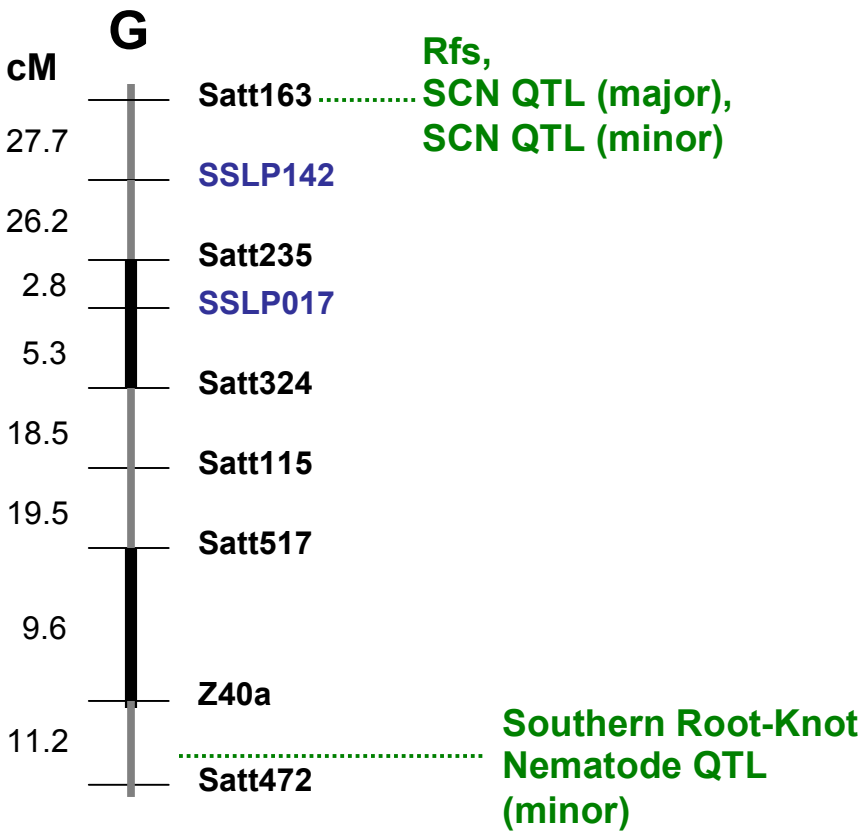


Figure 1 continued Genetic linkage map for soybean based on published framework (Cregan et al., 1999), SSLP, and RGA markers.

QTL If literature indicates the QTL is responsible for 30% or more of phenotypic variation then it is considered major, less than 30% is considered minor.

Rfs = Sudden death syndrome resistance gene

Abbreviations: (SCN) soybean cyst nematode

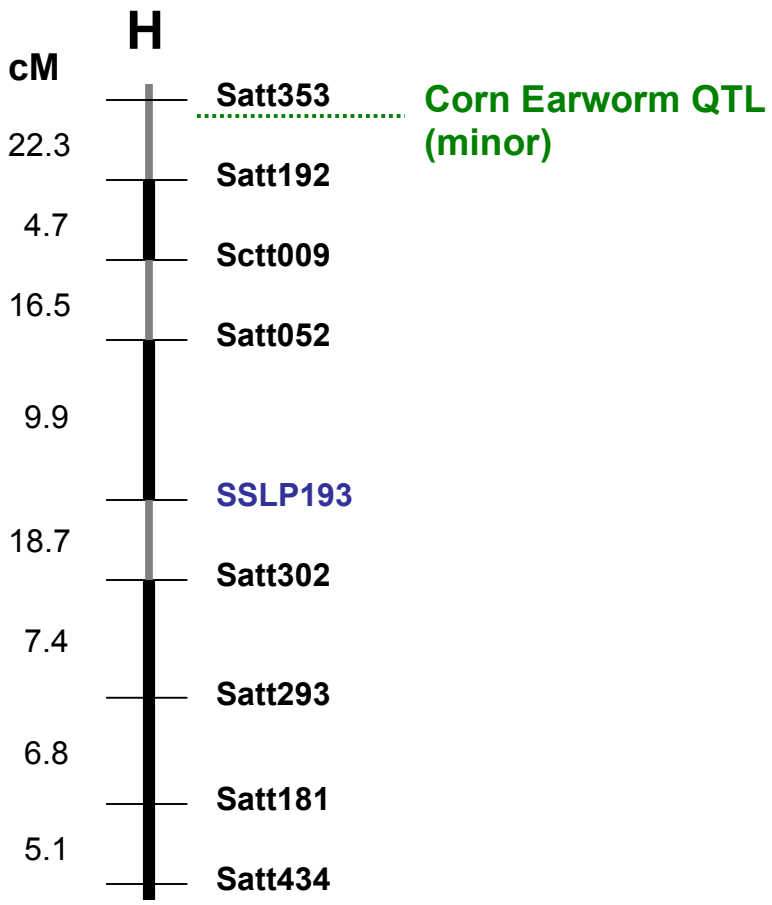


Figure 1 continued Genetic linkage map for soybean based on published framework (Cregan et al., 1999), SSLP, and RGA markers.

QTL If literature indicates the QTL is responsible for 30% or more of phenotypic variation then it is considered major, less than 30% is considered minor.

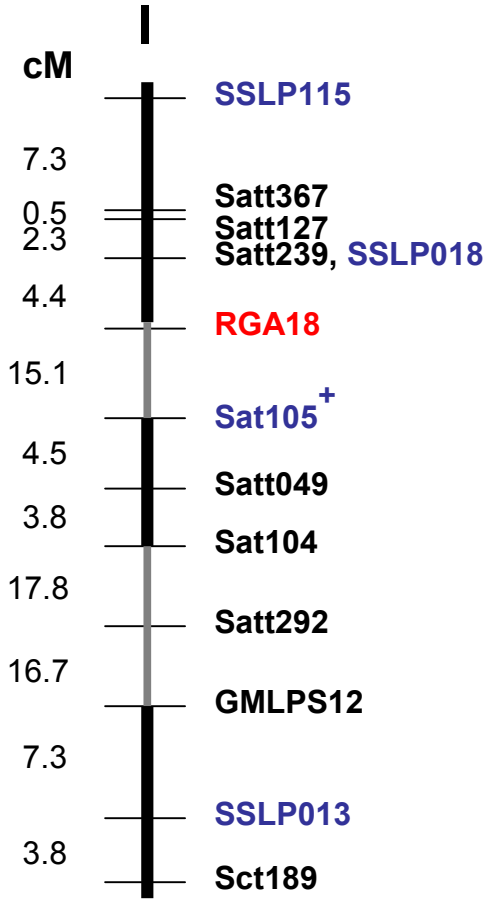


Figure 1 continued Genetic linkage map for soybean based on published framework (Cregan et al., 1999), SSLP, and RGA markers.

+ The marker has dual roles. It is a published framework marker; however, it is highlighted blue because its sequence information corresponds to a polymorphic site that was identified in one of our disease-related ESTs. Thus, it also represents an additional SSLP.

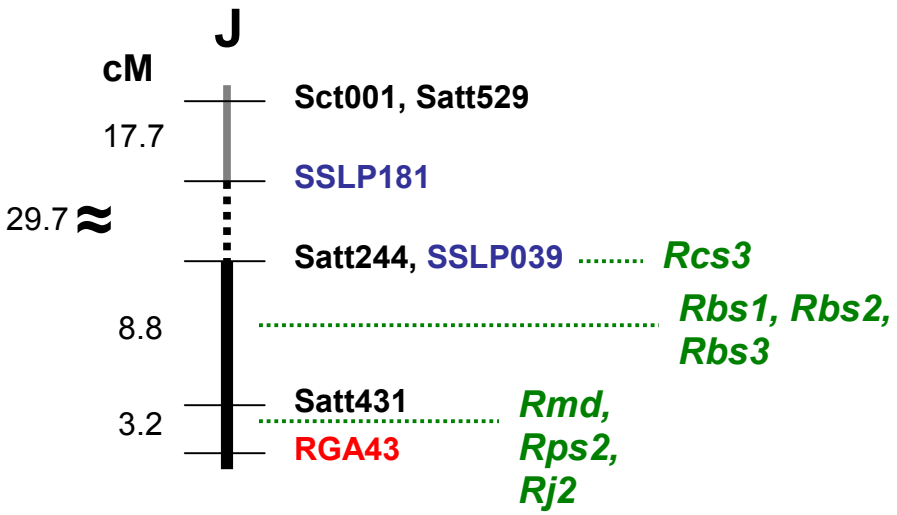


Figure 1 continued Genetic linkage map for soybean based on published framework (Cregan et al., 1999), SSLP, and RGA markers.

≈ Estimation for the gap distance is given to the left of the symbol. Estimations are based on markers location in several different mapping populations (Cregan et al., 1999).

Rcs3 = Frogeye leaf spot resistance gene
Rbs1, Rbs2, Rbs3 = Brown stem rot resistance genes
Rmd = Powdery mildew resistance gene
Rps2 = Phytopthora resistance gene
Rj2 = Nodulation resistance gene

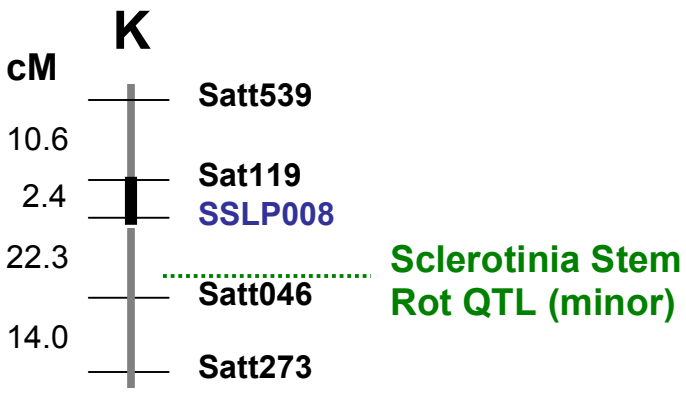


Figure 1 continued Genetic linkage map for soybean based on published framework (Cregan et al., 1999), SSLP, and RGA markers.

QTL If literature indicates the QTL is responsible for 30% or more of phenotypic variation then it is considered major, less than 30% is considered minor.

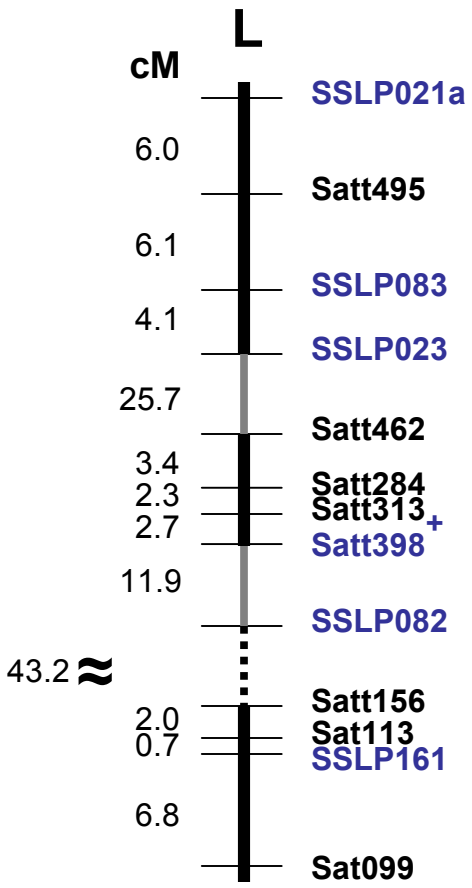


Figure 1 continued Genetic linkage map for soybean based on published framework (Cregan et al., 1999), SSLP, and RGA markers.

≈ Estimation for the gap distance is given to the left of the symbol. Estimations are based on markers location in several different mapping populations (Cregan et al., 1999).

+ The marker has dual roles. It is a published framework marker; however, it is highlighted blue because its sequence information corresponds to a polymorphic site that was identified in one of our disease-related ESTs. Thus, it also represents an additional SSLP.

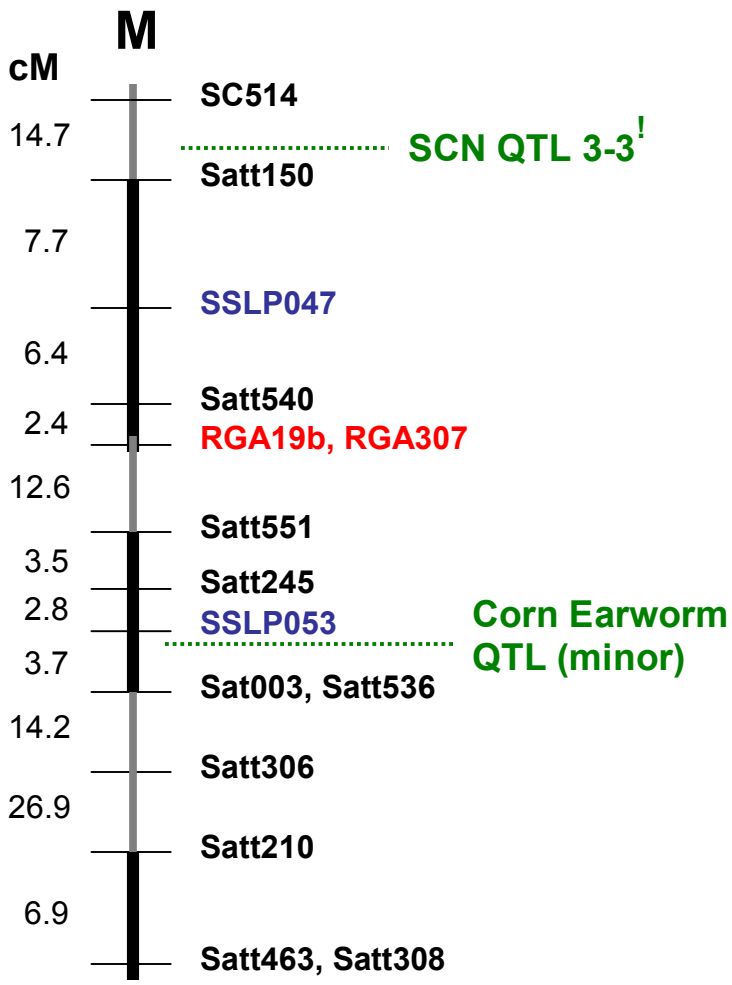


Figure 1 continued Genetic linkage map for soybean based on published framework (Cregan et al., 1999), SSLP, and RGA markers.

! QTL name as reported by Soybase (Shoemaker et al., 2003).

QTL If literature indicates the QTL is responsible for 30% or more of phenotypic variation then it is considered major, less than 30% is considered minor.

Abbreviations: (SCN) soybean cyst nematode

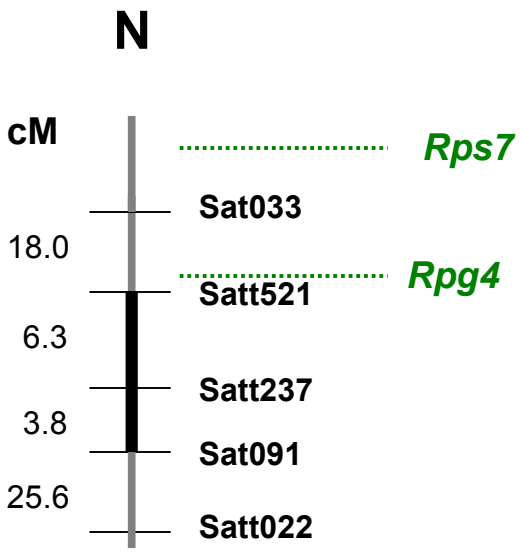


Figure 1 continued Genetic linkage map for soybean based on published framework (Cregan et al., 1999), SSLP, and RGA markers.

Rps7 = Phytopthora resistance gene
Rpg4 = Bacterial Blight resistance gene

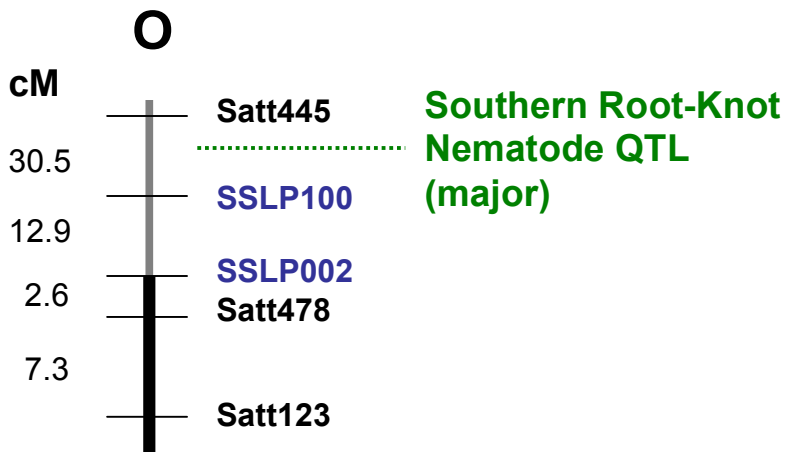


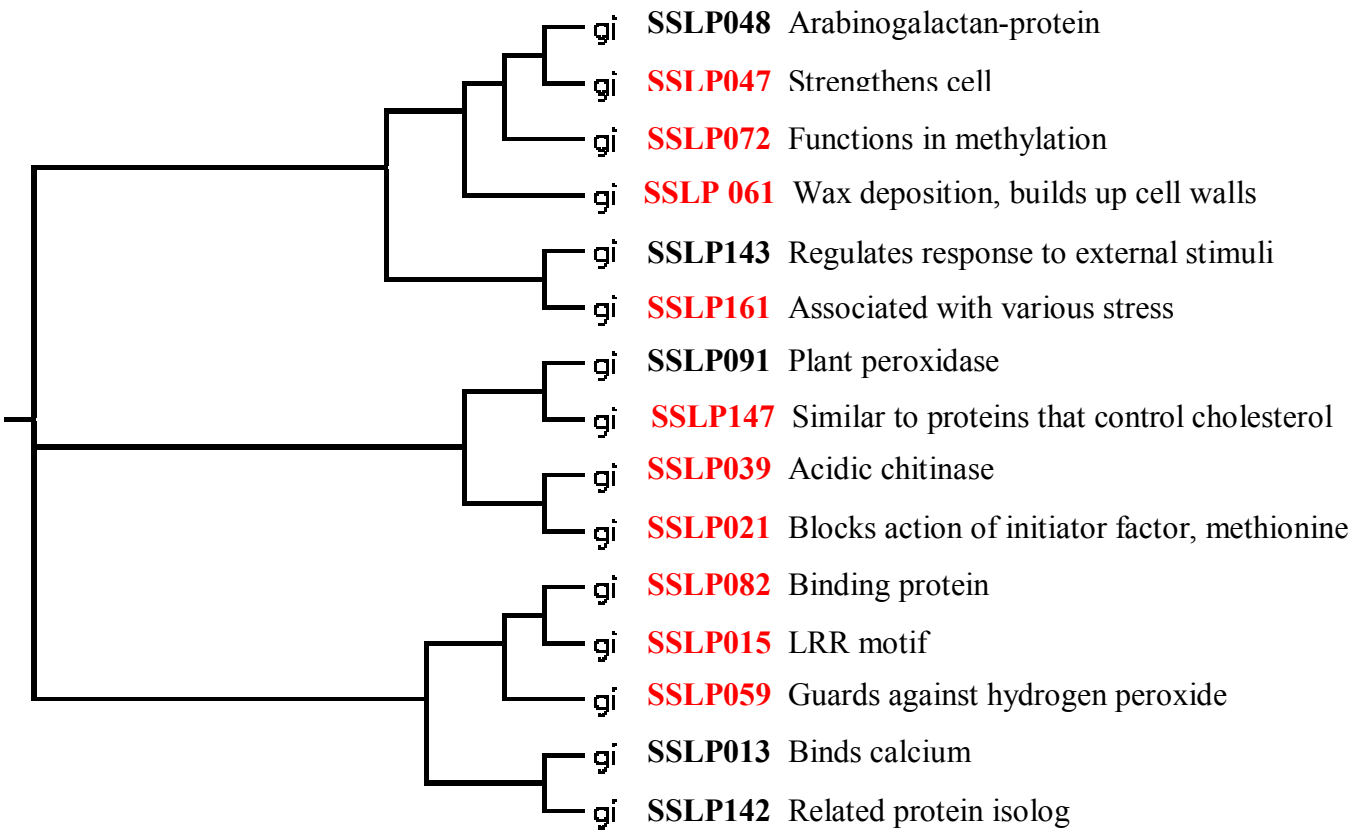
Figure 1 continued Genetic linkage map for soybean based on published framework (Cregan et al., 1999), SSLP, and RGA markers.

QTL If literature indicates the QTL is responsible for 30% or more of phenotypic variation then it is considered major, less than 30% is considered minor.

Figure 1 continued Genetic linkage map for soybean based on published framework (Cregan et al., 1999), SSLP, and RGA markers.

The unknown group did not contain any of the published markers and could not be included on the map with its appropriate linkage group.

The unlinked published primers are located near the tips of chromosomes and have distances that are far from the primers I have mapped in their respective linkage group.



*(gi) is followed by the ESTs' marker identification. The best description that could be determined for the ESTs function is given to the right of the SSLP ID. SSLPs highlighted in red were determined to be disease related.

Figure 2. Phylogenetic Analysis of 15 ESTs Characterized for a Putative Function. Three main groups are shown in the figure by the lines that branch out from the beginning. Within each group, sequences that are most related have the fewest branches between them. Related sequences have less evolutionary divergence, and should be involved in similar functions. This information can provide insights about specific resistance pathways of the ESTs. Also, ESTs that are better understood can be used to characterize less studied ESTs.

Chapter IV. **Summary and Future Prospects**

The improvement of soybeans as a crop can benefit tremendously from the improvement of molecular marker techniques. The key is to develop high through-put markers that are easily reproducible, transferable, and maximize the informativeness of available sequence information. The focus of this study was to demonstrate new marker strategies that will lead to improved disease resistance in soybeans. RGAs and SSLPs originating from disease-related ESTs were exploited to produce a set of markers directed towards saturating and tagging resistance genes.

A set of disease-related ESTs were scanned using SSLP technology to develop 202 PCR markers. Markers were also constructed to amplify the entire sequence of 22 previously identified RGAs. The markers were used to perform a diversity analysis on twelve soybean lines. Both types of markers were able to detect a high level of polymorphism in the soybean lines tested. The markers, along with SSR markers, were then used to construct a soybean linkage map. The SSR markers were used to relate our markers with the chromosomal locations of published disease resistance genes and QTLs. In the analysis, 16 resistance loci and 3 resistance QTLs were found to be associated with our markers. Four RGAs were related to resistance loci and all of them were mapped near SSLPs, which indicates that the markers could be part of a functional resistance gene.

The beauty of the markers used in this study is that they can be applied in a number of ways. Gene tagging for marker assisted selection in resistance breeding and saturating the position of resistance genes for cloning and transformation purposes are two areas that warrant further attention. The tip of molecular linkage group J would be an excellent place to begin. RGA43 and SSLP039 were mapped to an area on this linkage group that is scattered with several resistance loci. *Rsv4*, an important SMV resistance gene, is in the vicinity of SSLP090. This observation provides researchers with an additional marker to identify the gene. High resolution mapping has been successfully demonstrated by the work of Gore et al (2002) for *Rsv1*, another potyvirus gene in soybean.

The nature of ESTs and RGAs makes our markers excellent choices for the candidate gene approach. The idea is to design markers from sequence information that is believed to code for the gene of interest. Thus, the markers are generated from regions within the gene and can pinpoint its location precisely. Several studies have illustrated how effective RGAs can be applied in this manner, Gore et al. (2002) in soybean, Mago et al. (1999) in rice, and Collins et al. (2001) in barley. Since ESTs are designed from functional genes, it should be encouraging to also use these markers for this approach.

Any one of our SSLPs or RGAs could be applied to the candidate gene approach. It would probably be best to begin with markers that appear to be associated with resistance loci. However, markers not associated with any known resistance gene should not be ruled out. The highest number of SSLPs mapped to molecular linkage group D1a and L. D1a has QTLs for resistance to SCN, but no known resistance has been assigned to MLG L yet.

This study has provided a set of markers that can be used in a number of ways. SNP markers can be designed from SSLPs and applied for high throughput screening. SSLPs that were assigned putative functions related to disease resistance would be a good starting point for identifying markers that could be used by other researchers in disease studies. Further phylogenetic analysis could be performed on all SSLPs to identify more

markers related to the ones believed to be involved in resistance pathways. This work is the beginning for many other studies in disease resistance of soybeans. We have only mapped the location of 48 SSLPs. Identifying and mapping additional polymorphic SSLPs in our mapping population, along with any new RGAs, will provide more disease-related markers. In addition, it will lead to an improved soybean map that better represents the 20 MLGs in soybean by filling in the existing gaps of the map illustrated in Figure 1 of Chapter III. Our map construction was based on a subset of 114 RILs out of the 300 available in our laboratory. Therefore, our mapping population can be enlarged for high resolution mapping purposes and for segregation analysis of our markers with the known disease resistance genes.

References

Collins, N., Park, R., Spielmeyer, W., Ellis, J., and Pryor, A.J. 2001. Resistance gene analogs in barley and their relationship to rust resistance genes. Genome **44**: 375-381.

Gore, M.A., Hayes, A.J., Jeong, S.C., Yue Y.G., Buss, G.R. and Saghia Maroof, M.A. 2002. Mapping tightly linked genes controlling potyvirus infection at the *Rsv1* and *Rpv1* region in soybean. Genome. **45**: 592-599.

Mago, R., Nair, S., and Mohan, M. 1999. Resistance gene analogues from rice: cloning, sequencing and mapping. Theor. Appl. Genet. **99**: 50-57.

Michael Jason Godwin

Michael Jason Godwin was born January 19, 1978 in Virginia Beach, Virginia. Mr. Godwin graduated from Floyd E. Kellam High School in June 1996. In August 1996, he began college at Virginia Polytechnic Institute and State University, Blacksburg, Virginia, with an undecided major. After his second year of college, Mr. Godwin selected dual majors in Psychology and the Biotechnology concentration of Crop and Soil and Environmental Sciences. He participated with Dr. Russell Jones of the Psychology department in undergraduate research for children who suffered from posttraumatic stress disorder. Mr. Godwin graduated with his B. A. in May of 2001, and began graduate school in Crop and Soil Environmental Sciences, with a concentration of Genomics and Molecular Biology, during the last part of August, 2001. While beginning his graduate work he was allowed to finish his B. S., and completed it in December 2001. Upon completion of his M. S., Mr. Godwin will spend the remainder of his days surfing.