**Characterization of the soybean genome in regions surrounding two loci for resistance to soybean mosaic virus**

by

Alec J. Hayes

Dissertation submitted to the faculty of the
Virginia Polytechnic Institute and State University
in partial fulfillment of the requirements for the degree of
DOCTOR OF PHILOSOPHY
in
CROP AND SOIL ENVIRONMENTAL SCIENCES

_____
Dr. M. A. Saghai Maroof (chair)

_____ _____
Dr. Carol Wilkinson Dr. Glenn Buss

_____ _____
Dr. Sue Tolin Dr. Richard Veilleux

December 1998

Blacksburg, Virginia

**Characterization of the soybean genome in regions surrounding two loci for resistance to soybean mosaic virus**

by

Alec J. Hayes

(ABSTRACT)

Soybean mosaic virus (SMV), has been the cause of numerous and often devastating disease epidemics, causing reduction in both the quality and quantity of soybeans worldwide. Two important genes for resistance to SMV are *Rsv*1 and *Rsv*4. Alleles at the *Rsv*1 locus have been shown to control resistance to all but the most virulent strain of SMV. This locus has been mapped previously to the soybean F linkage group. *Rsv*4 is an SMV resistance locus independent of *Rsv*1 and confers resistance to all strains of SMV. This locus has not been mapped previously. The purpose of this study is to investigate the two genomic regions that contain these vitally important resistance genes.

A population of 281 $F_2$ individuals that had previously been genotyped for reaction to SMV was evaluated in a mapping study which combined bulk segregant analysis with Amplified Fragment Length Polymorphism (AFLP). A *Rsv*4-linked marker, R4-1, was identified that mapped to soybean linkage group D1b using a reference mapping population. More than 40 markers were mapped in the *Rsv*4 segregating population including eleven markers surrounding *Rsv*4. This

will provide the necessary framework for the fine mapping of this important genetic locus.

Previous work has located *Rsv*1 to a genomic region containing several important resistance genes including *Rps*3, *Rpg*1, and *Rpv*. An RFLP probe, NBS5, whose sequence closely resembles that of several cloned plant disease resistance genes has been mapped to this chromosomal region. The efficacy of using this sequence to identify potential disease resistance genes was assessed by screening a cDNA library to uncover a candidate disease resistance gene which corresponds to this NBS5 sequence. Two related sequence classes were identified that correspond to NBS5. Interestingly, one class corresponds to a full length gene closely resembling other previously cloned disease resistance genes offering evidence that this NBS5-derived clone is a candidate disease resistance gene.

A new marker technique was developed by combining the speed and efficiency of AFLP with DNA sequence information from cloned disease resistance genes. Using this strategy, three new markers tightly linked to *Rsv*1 were identified. One of these markers, which maps 0.6 cM away from *Rsv*1, has motifs consistent with other cloned disease resistance genes, providing evidence that this approach is an efficient

method for targeting genomic regions where disease resistance genes are located.

IN LOVING MEMORY OF LAREINE

**Acknowledgment**

In the course of completing my dissertation study I have had the pleasure to interact and learn from numerous talented and skillful individuals, without whom completion of this work would not have been possible. I am particularly grateful to my advisor, Dr. M. A. Saghai Maroof, whose knowledge and insight I greatly respect and whose ambition and success have been an inspiration. In addition I am indebted to my committee members, Dr. Carol Wilkinson, Dr. Glenn Buss, Dr. Sue Tolin and Dr. Richard Veilleux for their expertise both in guiding my graduate study, and in preparation of this dissertation.

There are numerous colleagues who have contributed directly and indirectly to this work and whose efforts are greatly appreciated. They include Dr. Ruslan Biyashev, Dr. Yong Yue, Dr. P. Jeff Maughan, Dr. Guo-rong Ma, and Dr. Pengyin Chen. Also I am grateful to have been able to work with such a large group of both gifted and amicable scientists who are too numerous to list. I would like to make mention of Mine Çiçek and Michael Gore, who have been very close friends throughout my tenure.

Finally, I owe a tremendous debt of gratitude to my supportive and loving family. They include my parents, Harlan and Betty; my two brothers, Mark and Brian; and their wives, Kim and Wanda. And to all the nieces and nephews, Amber, Austin, Zach, Brendan, and Megan, I love you guys.

**Table of Contents**

**List of Tables**

**List of Figures**

# CHAPTER I.

## Introduction

**Genetic Mapping in Soybean**

Since the rediscovery of Mendel's "Laws of Inheritance" in the early 20th century, the steady progression of genetic understanding as it relates to crop breeding has been unceasing. One of the more important concepts, genetic linkage, has been an especially powerful tool for the plant breeder. Genetic linkage is the principle that genes which are located close together on a chromosome are less likely to be recombined by a crossover event during meiosis than genes located far apart on a chromosome or on different chromosomes. Traits that are simply inherited and differentially expressed by a set of alleles at a locus can be scored in a population segregating for this trait. These qualitative traits can then be compared to other similarly obtained qualitative traits to determine if the two traits assort independently or if they are genetically linked. Any group of linked genetic loci (called a linkage group) correspond to all or a portion of a single chromosome.

In the early soybean breeding research of the late-1930s and 1940s, workers established genetic linkage maps that included traits such as seed color, flower color, leaf shape, and pubescence-type. In addition simply inherited traits of agronomic importance, such as pathogen resistance could be mapped on these linkage groups. Those

morphological markers that were determined to be closely linked to important genes such as disease resistance were of the greatest value, because the breeder could select for the morphological marker as a reasonably accurate indicator of the disease response genotype. In practice, this type of favorable linkage relationship is more the exception than the rule, because morphological markers are sparsely scattered on the soybean linkage map and often are not closely associated with any trait of interest.

**Molecular Markers**

It wasn't until the late 1970s that researchers began to explore new types of genetic markers to expand the soybean genetic map. Collectively termed molecular markers, these are detectable protein or nucleic acid variations that are simply inherited and segregate according to Mendel's laws. These new markers are generated in the laboratory using tools developed by biochemists and microbiologists.

The first of these markers to be used to establish genetic relationships were the isozyme markers. These were used largely in the 1970s and 1980s (Beremand, 1975). By analyzing the migration of selected enzymes, electrophoresed on a matrix such as a starch or polyacrylamide gel, researchers could detect slight differences in allelic

enzymes, or allozymes. The inheritance of these allozymes could be followed in a segregating population. For example, Rennie et al. (1987) reported a linkage relationship between soybean response to the pathogen *Phytophthora megasperma* f. sp. *glycinea* and the gel banding pattern of isocitrate dehydrogenase allozymes.

Allozyme mapping was hampered tremendously because of several factors. First there are very few enzymes identified which show detectable polymorphism, and thus the likelihood of finding a linkage relationship is very low. In addition, the expression of certain proteins throughout the life-cycle of the soybean and among tissue types could be highly variable, in addition to variability introduced by environmental conditions (Adams et al., 1981; Hanks et al., 1983). This adversely complicates the detection of some of the various isozymes in soybean.

**Restriction Fragment Length Polymorphism**

The evolution of molecular markers as a preeminent tool for soybean geneticists began in the late 1980s with the emergence of Southern blotting, which led to the detection of DNA polymorphisms, by Restriction Fragment Length Polymorphism (RFLP) (Apuya et al., 1988; Botstein et al., 1980). Detecting polymorphisms using DNA is a more powerful

genetic marker tool, because DNA is less variable across tissue types and throughout the life cycle of the plant than are proteins. RFLP is a technology whereby genomic DNA is broken down into short nucleotide chains by restriction enzymes that cut the DNA at very specific nucleotide sequences. Restriction enzymes that are routinely used for RFLP recognize specific palindromic sequences of between four and six base pairs (bp). This results in the genome being cut up into chains of nucleotides ranging from about 200 bp to about 20 kilobases (kb). Restriction-digested genomic DNA is run on an agarose gel that separates the DNA by size. The DNA is then transferred to a nylon membrane which can be hybridized using labeled DNA probes. Polymorphisms are detected when the probe hybridizes in at least one of several lines to a different size fragment. Polymorphisms typically arise from point mutations at the restriction enzyme cut site or through insertion or deletion of DNA.

RFLPs have been instrumental in developing the first saturated maps of soybean. Keim et al. (1990) reported the first linkage map of soybean, which was constructed based on linkage relationships of 150 RFLP markers in an interspecific population, *Glycine max* X *Glycine soja*. In this original work, the researchers were able to identify 26

linkage groups corresponding to the 20 chromosomes of soybean, far fewer linkage groups than had been previously identified. In addition, because these markers covered large regions of the soybean genome, the researchers were able to predict the genetic location of several major quantitative trait loci controlling important agronomic traits. This first RFLP map , which is known as the USDA-ARS-ISU map of soybean, has been combined with the classical genetic map, forming a reference map to which other soybean linkage maps are compared (Shoemaker and Specht, 1995). The linkage groups assigned to this map (A-W) are used by soybean breeders and molecular geneticists to describe the location of particular genes of interest. A database has been set up to list all the markers mapped in this population so that it may be used as a resource for obtaining markers that correspond to a particular genetic locus (http://probe.nalusda.gov).

RFLPs have become the standard genetic marker for soybean gene mapping because they can be generated fairly easily, and are highly repeatable from lab to lab. There are drawbacks to RFLP's as a genetic marker, however. First of all, the level of polymorphism detected by RFLP is fairly low among the cultivated and wild accessions of *G. max* and *G. soja* (Roth et al., 1989). Keim et al. (1989) reported

that among 58 diverse accessions of *G. max* and *G. soja*, 15 of 17 markers showed only two allele-types and the remaining two markers showed only three allele types. In addition 6 of the 17 markers showed an alternate allele in only one or two of the 58 lines. This observation suggests that the probability of detecting polymorphism in a given mapping population with a given probe is fairly low, particularly if both parents are elite lines and/or commercial cultivars. In fact in a related study of 38 adapted soybean lines and 128 RFLP marker loci, Keim et al. (1992) estimated that only one in five marker loci would be informative for a given pair of individuals. A similar study by Skorupska et al. (1993), testing 108 lines with 83 probes, reported similar results.

In addition to its low level of polymorphism, RFLP mapping of soybean is hampered by the complexity of the soybean genome. Soybean is postulated to be an ancient tetraploid and as such it contains significant regions of homoeology within its genome (Hymowitz and Singh, 1987; Lohnes et al., 1997). Homoeologous genomic regions are large stretches on two or more chromosomes, which are not identical, but share significant areas of highly conserved sequence. This means that a particular soybean probe will likely hybridize to more than one restriction fragment on a

Southern blot and often times will map to two or more different loci. This can complicate the association of a marker to the appropriate linkage group as it relates to a gene or locus of interest. In order to deal with this problem, researchers are careful to note the enzyme used and the sizes of the polymorphic bands mapped in order to avoid confusion. This information is not useful, however, should a researcher need to use a different restriction enzyme to detect a mappable polymorphism.

**PCR-based Molecular Markers**

In recent years, several different types of DNA markers have come into common use for mapping in soybean. Three of these molecular markers in common use are microsatellites or simple sequence repeats (SSR), random amplified polymorphic DNA (RAPD), and amplified fragment length polymorphism (AFLP). All of these markers are generated using the polymerase chain reaction (PCR). PCR is a method by which small DNA sequences (usually between 100 bp and 5 kb) can be amplified exponentially by thermocycling using a heat stable polymerase. These PCR-based markers address some of the concerns surrounding RFLPs.

SSR is a type of genetic marker that detects polymorphisms that result from differing copy number of a

small nucleotide repeat, of between 1 and 5 bp (Weber, 1990). Specific primer sequences are designed that flank a known small nucleotide repeat. The repeat is amplified by PCR and visualized on an agarose or polyacrylamide gel, in order to detect size polymorphisms resulting from the number of repeats. Akaya et al. (1992) reported polymorphism of AT and ATT repeat sequences in soybean. These were generated by designing primers and amplifying SSRs of sequences obtained from the Genbank database. They noted three markers that gave 6, 7, and 8 alleles per locus in a population study of 43 soybean genotypes. In numerous studies following this it has been noted that SSRs are a powerful genetic marker in soybean because they are highly polymorphic, they are found throughout the genome, they are easy to run once they have been identified, and they are almost always detected as a single locus marker (Diwan and Cregan, 1997; Doldi et al., 1997; Rongwen et al., 1995). The major drawback to microsattelites is that tremendous time and effort is required to locate the presence of an SSR in the genome.

Recently Cregan et al. (1998) reported a comprehensive map of soybean in which they mapped 443 SSR loci in at least one of four soybean mapping populations including the USDA-ARS-ISU soybean map. By using the computer program JOINMAP,

they were able to generate a consensus map that included information from markers mapped in all of these populations. In this way they were able to report 20 linkage groups in soybean corresponding to the 20 chromosomes. Because SSRs are single locus markers, this is the most informative molecular map thus far published in soybean.

RAPD is another PCR-based marker that has been used extensively (Martin et al., 1991; Williams et al., 1990). These markers are generated from genomic DNA template, using low stringency PCR and a single primer that is usually ten nucleotides in length. The idea is that by random chance, there will be several sites on the genome that contain two conversely oriented sites that are complementary to the random primer. Differential amplification of a product between two lines, usually presence/absence, results in a polymorphism being generated. RAPDs have not been used as markers for mapping as extensively in soybean as they have in some other species.

The major appeal of RAPDs as a marker technology is that they can be generated at a minimal cost to the researcher because a minimal amount of resources and genomic information are required up front. However, some studies have questioned the repeatability of RAPDs for certain applications (Chen et al., 1997). Specifically, these

markers are generated under low-stringency conditions. Minor changes in the PCR reaction, particularly from lab to lab, can alter the resulting banding pattern, calling into question the repeatability of some of these markers.

Recently, researchers have begun to deal with the repeatability issue by first identifying a RAPD band of interest and then isolating and sequencing this band to design sequence specific primers which reliably amplify the marker. These are referred to as sequence characterized amplified region (SCAR) markers and were first proposed by Paran and Michelmore (1993).

In recent years, there has been wide acceptance of AFLP as an important PCR-based marker system (Vos et al., 1995). AFLP is a relatively new type of molecular marker that allows researchers to screen large numbers of loci in a short period of time. To generate AFLPs, genomic DNA is double digested with a four- and six-cutter restriction enzyme. Specific nucleotide adaptors are then ligated to the restriction ends by DNA ligase. A subset of the generated fragments are amplified using adaptor-specific primers that contain three selective nucleotides at the 3' end of the primer. The six-cutter-specific primer is end-labeled using $\gamma^{32}$P-ATP and polynucleotide kinase. The PCR products are then run on a polyacrylamide gel and visualized

by autoradiography. Numerous bands (~40-60) are generated from a single run with AFLP.

Maughan et al. (1996) performed an AFLP study, evaluating 15 AFLP primer pairs for polymorphism in 12 *Glycine max* and 11 *Glycine soja* accessions. Of the 759 fragments detected, 274 (36%) were polymorphic between at least two accessions. While this level of polymorphism can not be expected to be seen in an intraspecific cross of closely related cultivars, the data suggest that AFLP is a powerful and efficient marker strategy available for detecting polymorphism among lines.

Keim et al. (1997) have recently published an AFLP-based linkage map in soybean. The map consists of 28 linkage groups covering 3441 cM. Six-hundred-fifty AFLP markers were mapped. The map contains a marker density of 1 per 4cM, but there are large gaps between markers in certain regions of the genome. This suggests that AFLP markers may not be as ubiquitous as once postulated. It is important to note however that this map was generated using only 42 recombinant inbred lines (RILs). While mapping with RILs increases the level of recombination relative to $F_2$ mapping, there are still relatively few segregating lines, lending some bias toward clustering of markers.

**Cloning Disease Resistance Genes**

The decade of the 1990s has witnessed a tremendous leap in the understanding of plant disease resistance with the cloning of several genes that directly trigger defense response. In 1992, Johal and Briggs cloned the Hm1 gene from maize (*Zea mays* L.). This disease resistance gene protects the maize plant from deleterious effects of the fungus, *Cochliobolus carbonum*, by neutralizing a toxin produced by the invading organism. While the cloning of this gene is certainly important, it is largely restricted to a single host-pathogen relationship and can not be applied to other plant-pathogen interactions.

A more common host-pathogen resistance complex is the gene-for-gene model of resistance, first described by Flor (1955). Nearly all plants studied, show some form of a gene-for-gene interaction. Essentially gene-for-gene means a single avirulence gene-product from the pathogen interacts with a single resistance gene product from the host to trigger a defense response that deters continued invasion and multiplication of the pathogen (Crute and Pink, 1996). The most common defense mechanism is the hypersensitive response in which the plant initiates localized cell death to deter the spread of the pathogen. The tomato (*Lycopersicon esculentum*) *Pto* and *P. syringae* pv. *tomato*

avrPto is one such gene-for-gene relationship that triggers a hypersensitive response (Ronald et al., 1992).

Ground-breaking work in the field of plant disease cloning began in 1993 when Martin et al. reported the successful map-based cloning of *Pto*, a single dominant gene in tomato which confers resistance to certain races of *Pseudomonas syringae* pv. *tomato*. This was the first cloning of a plant gene known to trigger localized defense responses. The *Pto* gene codes for a 321 amino acid, putative cytoplasmic protein with homology to protein kinases.

Following closely on the heels of the discovery of the *Pto* gene sequence was the successful cloning of two other resistance genes, *N* from tobacco (*Nicotiana tabacum*) (Whitham, 1994) and *Rps*2 from *Arabidopsis* (Bent et al., 1994). The *N* gene confers resistance to tobacco mosaic virus (TMV) and *Rps*2 provides resistance to the bacterial speck pathogen, *Pseudomonas syringae* pv. *tomato*. The elucidation of these two genes was important not only because they, like *Pto*, showed a gene-for-gene host pathogen relationship but because the two resistance genes themselves shared several regions of remarkable deduced amino acid conservation. This was especially important in light of the fact that these two genes are from very different organisms and confer resistance to widely differing pathogens.

Specifically, these gene products were shown to share marked homology in the N-terminal region, which exhibits a putative nucleotide binding site (NBS) region with conserved motifs resembling P-loop, kinase 2, and kinase 3a amino acid structures (Baker et al., 1997). Furthermore, these genes contain numerous other undefined conserved regions. In addition, both gene products contained a C-terminal leucine rich repeat (LRR) region. Neither of these genes shows any sequence similarity to the previously cloned *Pto* gene.

Several resistance genes from a small group of plants have been isolated since these original resistance genes were cloned. These include *Rpm*1 (Grant et al., 1995), *Rpp*5 (Parker et al., 1997), and *Rps*5 (Warren et al., 1998) from *Arabidopsis*, *Cf*-2 (Dixon et al., 1996), *Cf*-4 (Parniske et al., 1997), *Cf*-9 (Jones et al., 1994), *I*2C (Ori et al., 1997), and *Mi* (Milligan et al., 1998) from tomato, *Xa*1 (Yoshimura et al., 1998) and *Xa*21 (Song et al., 1995) from rice (*Oryza sativa*), *L*6 (Lawrence et al., 1995) and *M* (Anderson et al., 1997) from flax (*Linum usitatissimum*), and *Hs*1$^{pro-1}$ (Cai et al., 1997) from sugar beet (*Beta vulgaris*). In addition a gene proposed to be a strong candidate for *Cre*3 in wheat (*Triticum aestivum*) has been cloned (Lagudah et al., 1997). This gene has not been confirmed by transformation however.

These disease resistance genes can be divided into similarity groups based on deduced amino acid sequence (for a review see Baker et al., 1997). The three genes from tomato, *Cf*-2, *Cf*-4, and *Cf*-9, are similar in that they contain extensive regions of sequence identity and similarity including an LRR and a transmembrane domain. *Hs*1$^{pro-1}$ is also included in this group because it shares similar motifs with the *Cf* genes, although it does differ markedly from the *Cf* genes in overall amino acid sequence. Both *Xa*21 and *Pto* are classified uniquely individual because they differ significantly from all known resistance genes. *Xa*21 contains both a protein kinase motif and LRR. *Pto* contains a protein kinase domain. Subsequent studies have shown that *Pto* interacts directly with a second gene product, *Prf*, which contains an NBS and LRR (Salmeron et al., 1996). It is postulated that *Pto* is involved in pathogen recognition, while *Prf* actually triggers the cascade of events leading to a defense response.

By far the largest group of resistance genes belongs to the NBS-LRR group of genes. These include *N*, *Rps*2, *Rpm*1, *Rpp*5, *Rps*5, *I*2C, *Mi*, *Xa*1, *L*6, *M*, and the putative *Cre*3 gene. All of these genes share significant N-terminal conservation, including the NBS region that contains a conserved P-loop motif, GIxGMGGxGKTLA (x=any amino acid)

(Saraste et al., 1990), a kinase-2 region, K(R/K)α LαVLDDV(W/D) (α=aliphatic amino acid), and a putative kinase-3a region, GSRαIαTTR (Traut, 1994). In addition there are numerous regions of undefined conservation in the N-terminal region. Similarly all of these genes contain an LRR in their C-terminal region. The NBS region in these resistance genes is important because it is likely to facilitate signal transduction and thus is postulated to be involved in triggering the defense response. LRR's are not well understood but are thought to be involved in protein-protein interactions (Kobe and Deisenhofer, 1995) and thus could play a role in pathogen recognition (Baker et al., 1997).

**Homology-based Cloning of Resistance Gene-like Sequences**

To date, all these plant disease resistance genes have been cloned using either a map-based cloning strategy or a transposon-tagging strategy. These techniques are both tedious and tremendously labor intensive. The cloning of several of these NBS/LRR plant disease resistance genes with similarly conserved structural order has sparked tremendous interest in developing techniques that could identify other resistance genes that share similar structure but confer resistance to other pathogens. By using a homology-based

technique, one might be able to tremendously decrease both the time and resources required to clone a resistance gene of interest.

Yu et al. (1996) reported a PCR-based technique that was used in soybean to identify potential resistance-gene-like (RGL) sequences. The strategy consisted of designing degenerate primers from conserved sequences corresponding to two coding regions from the disease resistance genes *N* and *Rps*2. Specifically, these primers were designed based on the conserved P-loop and putative kinase-3a regions of the cloned genes. Using low-stringency polymerase chain reaction (PCR), they were able to amplify corresponding sequences of approximately 340 bp from soybean genomic DNA that contained similar conserved sequences. They determined that there are at least eleven different RGL classes present in soybean based on sequence information and hybridization patterns of the RGL sequences on a set of diagnostic Southern blots . The amplified sequences were then used as probes in a mapping population to determine their proximity to the location of known soybean disease resistance genes. Five of the eleven classes have been mapped to regions of the soybean genome where resistance genes have been identified.

At about the same time Kanazin et al. (1996) reported a similar study in soybean whereby they were able to employ a similar technique using slightly different primers to amplify RGL sequences. These workers also designed a degenerate primer corresponding to the P-loop region, but instead of using the kinase-3a as their reverse primer, they used a site farther downstream which corresponded to the highly conserved GLPLAL undefined motif. In their study they identified nine different RGL classes, six of which mapped near known disease resistance loci. Interestingly, of the nine RGL classes they identified, only three were closely related in sequence to the RGL sequences identified by Yu et al. (1996).

To date similar studies have been published by groups working with potato (*Solanum tuberosum*) (Leister et al., 1996), tomato (Ohmori et al., 1998), *Arabidopsis* (Aarts et al., 1998; Speulman et al., 1998), lettuce (*Lactuca sativa*) (Shen et al., 1998), sunflower (*Helianthus annuum*) (Gentzbittel et al., 1998) and in the grain crops, rice, wheat, barley (*Hordeum vulgare*), and maize (Collins et al., 1998; Leister et al. 1998; Seah et al., 1998). In all of these studies the researchers were able to identify RGL sequences that mapped very close to or cosegregated with resistance genes of interest. Other unpublished work has

turned up similar RGL sequences in oat (*Avena sativa*) and pine tree (*Pinus radiata*) (Cheng and Armstrong, unpublished; McCallum et al., unpublished).

In all, no less than 150 RGL sequences have been reported from various plant genomes. Comparison of these sequences along with the corresponding conserved sequence from previously cloned resistance genes shows that very different plant species, have very similar types of RGL sequences. For instance potato RGL clone, STU60080, shares 52.4% structural similarity with the lettuce RGL clone, AF017754 and 52.9% similarity with soybean RGL clone, NBS17. In addition, it is important to note that within a species RGL sequences can be highly divergent as well. For instance, *Arabidopsis* RGL sequence AF039377 is only 10.2% similar to another *Arabidopsis* RGL clone, AF039385, and is actually more similar to a group of potato RGL sequences. In general however, it is not surprising to notice that the stuctural tree of RGL sequences shows that most cluster into species specific groups.

Of course the importance of these RGL sequences is that they may represent small sequences within a larger gene that confers resistance to an important crop disease. In the works published thus far all have reported that at least several RGL sequences map near or cosegregate with a

resistance gene of interest. Certainly the potential exists to clone important genes directly using this technique. For instance, the recently cloned disease resistance gene, *Rps*5 of *Arabidopsis*, has been shown to contain an RGL sequence (Warren et al., 1998). The NBS portion of this gene was simultaneously and independently identified by Aarts et al. (1998), suggesting that it would have been possible to clone this gene simply by using the known RGL sequence.

Use of these sequences to tag a gene can be complicated however, because often times RGL sequences are members of a multi-gene family which cluster at a particular locus. Multiple copies of a similar RGL sequence may be repeated in a small genomic region, which would make it difficult to pinpoint which sequence, if any, actually represents the functional resistance gene. In addition, highly divergent RGL sequencess may be clustered very close to one another on the genome. For instance, Leister et al. (1998) identified an RGL sequence in rice, r4, that they physically linked to the Xa1 locus by isolating both on a single bacterial artificial chromosome (BAC). A group from Japan however had simultaneously cloned the Xa1 gene and shown that in fact it did not contain this RGL sequence, r4, but another somewhat divergent previously unidentified RGL sequence (Yoshimura et al., 1998).

**Resistance in Soybean to Soybean Mosaic Virus**

The most important viral disease of soybean worldwide is soybean mosaic virus disease caused by the soybean mosaic virus (SMV). The occurrence of this disease has been reported where ever soybeans are cultivated (Thottapilly and Rossel, 1987). Outbreaks of the disease, particularly in regions of Asia, have been blamed for severe yield losses (Irwin and Goodman, 1981). The disease has been reported throughout the U.S. soybean production area with recent yield losses being most severe in the lower Mississippi Delta (S. Tolin, personal communication). The upper South has also experienced significant yield losses as a result of SMV disease (Ren et al., 1997).

Symptoms of SMV infection on susceptible cultivars include typical mosaic-like dark and light green areas, as well as irregular areas of raised leaf tissue and downward curling of the leaf margins (Sinclair, 1982). General stunting of infected plants is also a common manifestation of SMV infection. A final important symptom of SMV infection is seed coat mottling which results in reduced value due to decreased seed quality (Quiniones et al., 1971).

As with most viral diseases in plants, efforts to control the spread of the pathogen primarily involves the

development and cultivation of resistant cultivars. Three genetic loci controlling resistance to SMV have been reported to date. *Rsv*1, first reported by Kihl and Hartwig (1979), is the most widely used resistance source in commercial soybean cultivars. Two additional SMV resistance loci, *Rsv*3 and *Rsv*4, have been reported (Ma et al., 1995; Tu and Buzzell, 1987).

Resistance at the *Rsv*1 locus has been reported from numerous sources. Different alleles at this locus have been reported including *Rsv*1-y from York, *Rsv*1-m from Marshall, *Rsv*1-k from Kwanggyo, *Rsv*1-t from Ogden, and *Rsv*1-n from PI486355 (Chen et al., 1991, Ma et al., 1995). All of these alleles can be distinguished by their reaction to the seven strains of SMV G1-G7, as distinguished by Cho and Goodman (1979). The most important of these *Rsv*1 alleles, derived from PI96983, confers resistance to all but the most virulent strain, G7, of SMV. This allele has been mapped in a population of 243 $F_2$ individuals of the cross PI96983 X Lee68, to a cluster of disease resistance genes on the F-linkage group, flanked by the RFLP markers, K644 and B212 (Yu et al., 1994). *Rsv*1 maps 0.2 cM from the RFLP marker R45 (Saghai Maroof, unpublished).

Two additional independently segregating SMV resistance loci are *Rsv*3 and *Rsv*4. *Rsv*3 from L29 confers resistance to

the more virulent SMV strains, G5-G7, but shows susceptibility to strains G1-G4. This gene was originally derived from the soybean cultivar Columbia (Tu and Buzzell, 1987). *Rsv*4 is a single dominant resistance gene derived from PI486355 (Ma et al., 1995). It confers resistance to all seven strain groups of SMV and unlike all other SMV resistance genes does not show any necrotic response. Map locations for the two genes, *Rsv*3 and *Rsv*4, have not been reported in the literature.

**Research Objectives**

The major focus of this research is to investigate the genomic regions surrounding the two SMV resistance genes *Rsv*1 and *Rsv*4. As stated earlier, *Rsv*1 maps to a resistance gene cluster on linkage group F. Around this cluster of genes, two RGL sequences, class b and j, have been identified, suggesting that the potential exists to identify resistance genes and other RGL sequences in this important chromosomal region. The chromosomal region where *Rsv*4 is located has not been reported, thus preliminary mapping work needs to be done to characterize the *Rsv*4 locus.

The specific objectives of this research are:

1) to combine AFLP and bulk segregant analysis strategies in a segregating $F_{2:3}$ population of D26 X Lee 68

(Ma, 1995) to locate the chromosomal region containing *Rsv*4 and to then map the *Rsv*4 region using known markers for that linkage group.

2) to expand on the NBS work begun by Dr. Yong Yu by screening a cDNA library from soybean using the class b NBS5 probe that maps to the *Rsv*1 region, in order to clone and characterize a full length disease resistance-like gene.

3) to develop and test a novel marker system that combines RGL homology-based strategies with AFLP technology to generate markers tightly linked to the *Rsv*1 gene.

**Literature Cited**

Aarts, M. G. M., B. L. Hekkert, E. B. Holub, J. L. Beynon, W. J. Stiekema, and A. Pereira. 1998. Identification of R-gene homologous DNA fragments genetically linked to disease resistance loci in *Arabidopsis thaliana*. Molecular Plant Microbe Interactions 11:251-258.

Adams, C.A., T. H. Broman, S. W. Norby, and R. W. Rinne. 1981. Occurrence of multiple forms of alpha amylase and absence of starch phosphorylase in soyabean seeds. Annals of Botany 48:895-903.

Anderson, P., G. Lawrence, B. Morrish, M. Ayliffe, J. Finnegan, and J. Ellis. 1997. Inactivation of the flax rust resistance gene M associated with loss of a repeated unit within the leucine-rich repeat coding region. Plant Cell 9:641-651.

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

Apuya, N. R., B. L. Frazier, P. Keim, E. J. Roth, and K. G. Lark. 1988. Restriction fragment length polymorphisms as genetic markers in soybean, *Glycine max* (L.) Merrill. Theoretical and Applied Genetics 75:889-901.

Baker, B., P. Zambryski, B. Staskawicz, and S.P. Dinesh-Kumar. 1997. Signaling in plant microbe interactions. Science 276:726-733.

Bent, A. F., B. N. Kunkel, D. Dahlbeck, K. L. Brown, R. Schmidt, J. Giraudat, J. Leung, and B. Staskawicz. 1994. Rps2 of *Arabidopsis* thaliana: a leucine-rich repeat class of plant disease resistance genes. Science 265:1856-1860.

Beremand, M. L. 1975. Soybean alcohol dehydrogenase: dimeric structure and subunit composition of multiple isozymes. Genetics 80:3.

Botstein, D., R. L. White, M. Skolnick, and R. W. Davis. 1980. Construction of a genetic linkage map in man using restriction fragment length polymorphism. American Journal of Human Genetics 32:314-331.

Cai, D., M. Kleine, S. Kifle, H. J. Harloff, N. N. Sandal, K. A. Marcker, R. M. Klein-Lankhorst, E. M. J. Salentijn, W.

Lange, W. J. Stiekema, U. Wyss, F. M. W. Grundler, and C. Jung. 1997. Positional cloning of a gene for nematode resistance in sugar beet. Science 275:832-834.

Chen, L., H. Kuo, M. Chen, K. Lai, S. Chen, H. Kuo, M. Chen, and K. Lai. 1997. Reproducibility of the differential amplification between leaf and root DNAs in soybean revealed by RAPD markers. Theoretical and Applied Genetics 95:1033-1043.

Chen, P., G. R. Buss, C. W. Roane, and S. A. Tolin. 1991. Allelism among genes for resistance to soybean mosaic virus in strain-differential soybean cultivars. Crop Science 31:305-309.

Cho, E. K., and R. M. Goodman. 1979. Strains of soybean mosaic virus. Classification based on virulence in resistant soybean cultivars. Phytopathology 69:467-470.

Collins, N. C., C. A. Webb, S. Seah, J. G. Ellis, S. H. Hulbert, and A. Pryor. 1998. The isolation and mapping of disease resistance gene analogs in maize. Molecular Plant Microbe Interactions 11:968-978.

Cregan, P. B., T. Jarvik, A. L. Bush, R. C. Shoemaker, K. G. Lark, A. L. Kahler, T. T. Van Toai, D. G. Lohnes, J. Chung, and J. E. Specht. 1998. An integrated genetic linkage map of soybean. Crop Science: (in press).

Crute, I. R., and D. A. C. Pink. 1996. Genetics and utilization of pathogen resistance in plants. Plant Cell 8:1747-1755.

Diwan, N and P. B. Cregan. 1997. Automated sizing of fluorescent-labeled simple sequence repeat (SSR) markers to assay genetic variation in soybean. Theoretical and Applied Genetics 95:723-733.

Dixon, M.S., D. A. Jones, J. S. Keddie, C. M. Thomas, K. Harrison, and J. D. Jones. 1996. The tomato Cf-2 disease resistance locus comprises two functional genes encoding leucine-rich repeat proteins. Cell 84:451-459 (1996)

Doldi M. L., J. Vollmann, and T. Lelley. 1997. Genetic diversity in soybean as determined by RAPD and microsatellite analysis. Plant Breeding 116:331-335.

Flor, A. H. 1955. Host-parasite interactions in flax rust - its genetics and other implications. Phytopathology 45:680-685.

Gentzbittel, L., S. Mouzeyar, S. Badaoui, E. Mestries, F. Vear, D. T. D. Labrouhe, and P. Nicolas. 1998. Cloning of molecular markers for disease resistance in sunflower, *Helianthus annuus*, L. Theoretical and Applied Genetics 96:519-525.

Grant, M. R., L. Godiard, E. Straube, T. Ashfield, J. Lewald, A. Sattler, R. W. Innes, and J. L. Dangl. 1995. Structure of the *Arabidopsis Rpm1* gene enabling dual specificity disease resistance. Science 269:843-846.

Hanks, J.F., K. Schubert, and N. E. Tolbert. 1983. Isolation and characterization of infected and uninfected cells from soybean nodules. Role of uninfected cells in ureide synthesis. Plant Physiology 71:869-873.

Hymowitz, T. and R. J. Singh. 1987. Taxonomy and Speciation. Agronomy 16:23-48.

Irwin, M. E. and R. M. Goodman. 1981. Ecology and control of soybean mosaic virus. pp. 181-220 *in* Maramorosch, K. and K. F. Harris (eds.) Plant diseases and vectors: ecology and epidemiology. Academic Press, New York.

Johal, G. S. and S. P. Briggs. 1992. Reductase activity encoded by the HM1 disease resistance gene in maize. Science 258:985-987.

Jones, D.A., C. M. Thomas, K. E. Hammond-Kosack, P. J. Balint-Kurti, and J. D. Jones. 1994. Isolation of the tomato Cf-9 gene for resistance to *Cladosporium fulvum* by transposon tagging. Science 266:789-793.

Kanazin, V., L. F. Marek, and R. C. Shoemaker. 1996. Resistance gene analogs are conserved and clustered in soybean. Proceedings of the National Academy of Sciences, USA 93:11746-11750.

Keim, P., W. Beavis, J. Schupp, and R. Freestone. 1992. Evaluation of soybean RFLP marker diversity in adapted germ plasm. Theoretical and Applied Genetics 85:205-212.

Keim, P., B. W. Diers, T. C. Olson, and R. C. Shoemaker. 1990. RFLP mapping in soybean: association between marker

loci and variation in quantative traits. Genetics 126:735-742.

Keim, P., J. M. Schupp, S. E. Travis, K. Clayton, and D. M. Webb. 1997. A high-density soybean genetic map based upon AFLP markers. Crop Science 37:537-543.

Keim, P., R. C. Shoemaker, and R. G. Palmer. 1989. Restriction fragment length polymorphism diversity in soybean. Theoretical and Applied Genetics 77:786-792.

Kihl, R. A. S., and E. E. Hartwig. 1979. Inheritance of reaction to soybean mosaic virus in soybeans. Crop Science 19:372-375.

Kobe, B. and J. Deisenhofer. 1995. A structural basis of the interactions betweeen leucine-rich repeats and protein ligands. Nature 374:183-185.

Lagudah, E. S., O. Moullet, and R. Appels. 1997. Map-based cloning of a gene sequence encoding a nucleotide-binding domain and a leucine-rich region at the *Cre*3 nematode resistance locus of wheat. Genome 40:659-665.

Lawrence, G. J., E. J. Finnegan, M. A. Ayliffe, and J. G. Ellis. 1995. The *L*6 gene for flax rust resistance is related to the *Arabidopsis* bacterial resistance gene *Rps*2 and the tobacco viral resistance gene N. Plant Cell 7:1195-1206.

Leister, D., A. Ballvora, F. Salamini, and C. Gephardt. 1996. A PCR-based approach for isolating pathogen resistance genes from potato with potential for wide application in plants. Nature Genetics 14:421-428.

Leister, D., J. Kurth, D. A. Laurie, M. Yano, T. Sasaki, K. Devos, A. Graner, and P. Schulze-Lefert. 1998. Rapid reorganization of resistance gene homologues in cereal genomes. Proceedings of the National Academy of Sciences, USA USA 95:370-375.

Lohnes, D. G., J. E. Specht, and P. B. Cregan. 1997. Evidence for homoeologous linkage groups in the soybean. Crop Science 37:254-257.

Ma, G. 1996. Genetic analysis of soybean reactions to soybean mosaic virus. PhD dissertation. Virginia Tech.

Ma, G., P. Chen, G. R. Buss, and S. A. Tolin. 1995. Genetic characteristics of two genes for resistance to soybean mosaic virus in PI486355 soybean. Theoretical and Applied Genetics 91:907-914.

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

Martin, G. B., J. Williams, and S. D. Tanksley. 1991. Rapid identification of markers linked to a *Pseudomonas* resistance gene in tomato by using random primers and near-isogenic lines. Proceedings of the National Academy of Sciences, USA 88:2336-2340.

Maughan, P. J., M. A. Saghai Maroof, G. R. Buss, and G. M. Huestis. 1996. Amplified fragment length polymorphism (AFLP) in soybean: species diversity, inheritance, and near-isogenic line analysis. Theoretical and Applied Genetics 93:392-401.

Milligan, S. B., J. Bodeau, J. Yaghoobi, I. Kaloshian, P. Zabel, and V. M. Williamson. 1998. The root knot nematode resistance gene *Mi* from tomato is a member of the leucine zipper, nucleotide binding, leucine-rich repeat family of plant genes. Plant Cell 10:1307-1319.

Ohmori, T., M. Murata, and F. Motoyoshi. 1998. Characterization of disease resistance gene-like sequences in near-isogenic lines of tomato. Theoretical and Applied Genetics 96:331-338.

Ori, N., Y. Eshed, I. Paran, G. Presting, D. Aviv, S. Tanksley, D. Zamir, and R. Fluhr. 1997. The *I*2C family from the wilt disease resistance locus *I*2 belongs to the nucleotide binding, leucine-rich repeat superfamily of plant resistance genes. Plant Cell 9:521-532.

Paran, I and R. W. Michelmore. 1993. Development of reliable PCR-based markers linked to downy mildew resistance genes in lettuce. Theoretical and Applied Genetics 85:985-993.

Parker, J. E., M. J. Coleman, V. Szabo, L. N. Frost, R. Schmidt, E. A. van der Biezen, T. Moores, C. Dean, M. J. Daniels, and J. D. G. Jones. 1997. The *Arabidopsis* downy mildew resistance gene *RPP*5 shares similarity to the Toll

and interleukin-1 receptors with *N* and *L*6. Plant Cell 9:879-894.

Parniske, M., K. E. Hammond-Kosack, C. Golstein, C. M. Thomas, D. A. Jones, K. Harrison, B. B. Wulff and J. D. Jones. 1997. Novel disease resistance specificities result from sequence exchange between tandemly repeated genes at the *Cf*-4/9 locus of tomato. Cell 91:821-832.

Quinones, S. S., J. M. Dunleavy, and J. W. Fisher. 1971. Performance of three soybean varieties inoculated with soybean mosaic virus and bean mottle virus. Crop Science 11:662-664.

Ren, Q., T. W. Pfeiffer, and S. A. Ghabrial. 1997. Soybean mosaic virus incidence level and infection time: Interaction effects on soybean. Crop Science 37:1706-1711.

Rennie, B. D., W. D. Beversdorf, and R. I. Buzzell. 1987. The potential use of isozymes as genetic markers for disease resistance: an example from the soybean/*Phytophthora* system. Canadian Journal of Plant Science 67:285.

Ronald, P. C., J. M. Salmeron, F. M. Carland, and B. J. Staskawicz. 1992. The cloned avirulence gene avr*Pto* induces disease resistance in tomato cultivars containing the *Pto* resistance gene. Journal of Bacteriology 174:1604-1611.

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

Roth, E. J., B. L. Frazier, N. R. Apuya, and K. G. Lark. 1989. Genetic variation in an inbred plant: variation in tissue cultures of soybean (*Glycine max* (L.) Merrill) Genetics 121:359-368.

Salmeron, J. M., G. E. D. Oldroyd, C. M. T. Rommens, S. R. Scofield, H. S. Kim, D. T. lavelle, D. Dahlbeck, and B. Stascawicz. 1996. Tomato *Prf* is a member of the leucine-rich repeat class of plant disease resistance genes and lies embedded within the Pto kinase gene cluster. Cell 86:123-133.

Saraste, M., P. R. Sibbald, and A. Whittinghofer. 1990. The P-loop - a common motif in ATP- and GTP-binding proteins. TIBS 15:430-434.

Seah, S., K. Sivasithamparam, A. Karakousis, and E. Lagudah. 1998. Cloning and characterisation of a family of disease resistance gene analogs from wheat and barley. Theoretical and Applied Genetics 97:937-945.

Shen, K. A., B. C. Meyers, M. N. I. Faridi, D. B. Chin, D. M. Stelly, and R. W. Michelmore. 1998. Resistance gene candidates identified by PCR with degenerate oligonucleotide primers map to clusters of resistance genes in lettuce. Molecular Plant Microbe Interactions 11:815-823.

Shoemaker, R. C. and J. E. Specht. 1995. Integration of the soybean molecular and classical genetic linkage groups. Crop Science 35: 436-446.

Sinclair, J. B. 1982. Compendium of soybean diseases. The American Phytopathology Society, St. Paul, Minn. 104pp.

Skorupska, H. T., R. C. Shoemaker, A. Warner, E. R. Shipe, and W. C. Bridges. 1993. Restriction fragment length polymorphism in soybean germplasm of the southern USA. Crop Science 33:1169-1176.

Song, W. Y., G. L. Wang, L. L. Chen, H. S. Kim, L. Y. Pi, T. Holsten, J. Gardner, B. Wang, W. X. Zhai, L. H. Zhu, C. Fauquet, and P. Ronald. 1995. A receptor kinase-like protein encoded by the rice disease resistance gene, *Xa*21. Science 270:1804-1806.

Speulman, E., D. Bouchez, E. B. Holub, and J. Beynon. 1998. Disease resistance gene homologs correlate with disease resistance loci of *Arabidopsis thaliana*. Plant Journal 14: 467-475

Thottapilly, G. and H. W. Rossel. 1987. Viruses affecting soybean *in* Singh, S. R., K. O. Rachie, and K. E. Dashiell (eds.). Soybeans for the tropics: research, production, utilization. John Wiley and Sons Ltd., Chichester, U.K. pp.53-68.

Traut, T. W. 1994. The functions and consensus motifs of nine types of peptide segments that form different types of nucleotide-binding sites. European Journal of Biochemistry 222:9-19.

Tu, J. C., and R. I. Buzzell. 1987. Stem-tip necrosis: a hypersensitive, temperature dependant, dominant gene reaction of soybean to infection by soybean mosaic virus. Canadian Journal of Plant Science 67:661-665.

Vos, P., R. Hogers, M. Bleeker, M. Reijans, T. van de Lee, M. Hornes, A. Frijters, J. Pot, J. Peleman, M. Kuiper, and M. Zabeau. 1995. AFLP: a new technique for DNA fingerprinting. Nucleic Acids Research 23:4407-4414.

Warren, R. F., A. Henk, P. Mowery, E. Holub, and R. W. Innes. 1998. A mutation within the leucine-rich repeat domain of the *Arabidopsis* disease resistance gene *Rps*5 partially suppresses multiple bacterial and downy mildew resistance genes. Plant Cell 10:1439-1452.

Weber, J. L. 1990. Informativeness of human (dC-dA)n.(dG-dT)n polymorphisms. Genomics 7:524-530.

Whitham, S., S. P. Dinesh-Kumar, D. Choi, R. Hehl, C. Corr, and B. Baker. 1994. The product of the tobacco mosaic virus resistance gene *N*: similarity to toll and the interleukin-1 receptor. Cell 78:1101-1115.

Williams, J.G.K., A. R. Kubelik, K. J. Livak, J. A. Rafalski, and S. V. Tingey. 1990. DNA polymorphisms amplified by arbitrary primers are useful as genetic markers. Nucleic Acids Research 18:6531-6535.

Yoshimura, S., U. Yamanouchi, Y. Katayose, S. Toki, Z. Wang, I. Kono, N. Kurata, M. Yano, N. Iwata, and T. Sasaki. 1998. Expression of *Xa*1, a bacterial blight-resistance gene in rice, is induced by bacterial inoculation. Proceedings of the National Academy of Sciences, USA 95:1663-1668.

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

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

# CHAPTER II

# MOLECULAR MARKER MAPPING OF *Rsv4*, A GENE CONFERRING RESISTANCE TO ALL STRAINS OF SOYBEAN MOSAIC VIRUS

**Abstract**

Several genes that confer resistance in soybean to soybean mosaic virus (SMV) have been identified. One of these resistance loci, *Rsv*4, confers resistance to all known strain groups of SMV. A population of 281 $F_2$ individuals from the cross of D26 (*Rsv*4) X Lee 68 (*rsv*4) was screened for disease response. Two sets of DNA from 12 to 15 $F_2$ individuals being either homozygous resistant or susceptible were pooled to produce two bulk resistant and two bulk susceptible DNA samples. Parents and bulks were screened with 101 AFLP primer pairs and linked polymorphisms were identified. A putative linked locus, amplified by the primers *Mse*-AAA and *Eco*-AAG, was converted to an RFLP and a single polymorphic band was mapped 5 cM away from *Rsv*4. The same band was mapped in a second population containing 300 mapped loci and it was determined that this band was located on the D1b linkage group. Subsequent mapping in the *Rsv*4 mapping population has shown that the *Rsv*4 locus is flanked by the microsattelite markers, Satt542 at 4.7 cM and Satt558 at 7.8 cM.

**Introduction**

The most important viral disease of soybean worldwide is caused by the soybean mosaic virus (SMV). The occurrence of this disease has been reported where ever soybeans are cultivated (Thottapilly and Rossel, 1987). Outbreaks of the disease, particularly in regions of Asia, have been blamed for severe yield losses (Irwin and Goodman, 1981). The disease has been reported throughout the U.S. soybean production area with recent yield losses being most severe in the lower Mississippi Delta (S. Tolin, personal communication). Incidence in the upper South have also been blamed for significant yield reduction (Ren et al., 1997). Efforts to control the incidence and spread of SMV revolve mainly around the development and utilization of cultivars possessing SMV resistance.

To date, three distinct genetic loci have been reported that confer resistance to SMV. The first locus to be identified, *Rsv*1 (Kihl and Hartwig, 1979), is the source of resistance in most commercially available cultivars. Resistance alleles at this locus have been effective against all but the most virulent strain, G7, of SMV. Yu et al.(1994) mapped the *Rsv*1 locus on the soybean F linkage group to a cluster of resistance genes flanked by the RFLP

markers pK644 and pB212. *Rsv*1 maps 0.2 cM from the RFLP marker, R45 (Saghai Maroof, unpublished).

A second SMV resistance locus, *Rsv*3, derived from the cultivar 'Columbia', was reported by Tu and Buzzel (1987). *Rsv*3 is a dominant resistance gene which is characterized by stem-tip necrosis following infection by strains G1-G4 of SMV. The more virulent strains G5-G7 are incapable of infecting lines possessing the *Rsv*3 gene. This locus has been mapped to the B linkage group of soybean, 0.9 cM from the RFLP locus, A519 (Kristipati, unpublished).

A third resistance locus was reported by Ma et al. (1995). This gene, designated *Rsv*4, confers resistance to all known strains of SMV. It is derived from line PI486355, which was shown to contain two resistance loci, one of which is not allelic at either the *Rsv*1 or *Rsv*3 locus. It is of interest to note that this resistance gene is completely dominant. Many SMV resistance alleles show systemic necrosis when present in the heterozygous state (Chen et al., 1994). The resistance allele at the *Rsv*4 locus shows resistance without necrosis in both the heterozygous and homozygous states.

A major focus of this research is to map the genetic location of *Rsv*4 on the soybean linkage map. By using Amplified Fragment Length Polymorphism (AFLP)(Vos et al.,

1995) and bulk segregant analysis (Michelmore et al., 1991) we have identified markers closely linked to *Rsv4*. AFLP is a PCR-based molecular marker that allows researchers to screen large numbers of loci in a short period of time. Specifically, genomic DNA is double digested with a four- and six- cutter restriction enzyme. Specific adaptors are then ligated to the restriction ends by DNA ligase. A subset of the generated fragments are amplified using adaptor specific primers that contain three selective nucleotides at the 3' end of the primer. The six-cutter-specific primer is end-labeled using $\gamma^{32}$P-ATP and polynucleotide kinase. The PCR products are then run on a polyacrylamide gel and visualized by autoradiography.

By combining AFLP with bulk segregant analysis, one can quickly identify markers closely linked to a gene of interest. Bulk segregant analysis (BSA), first proposed by Michelmore et al. (1991), is a strategy by which DNA from several $F_2$ individuals (typically 12 to 15) of a segregating population are pooled based on a selective criterion (in this case disease resistance/susceptibility). Marker genotypes for the disease susceptible bulk and disease resistant bulk are determined in addition to parental genotypes. Polymorphisms in which the bulked samples show the same marker genotype as their respective donor parent

represent closely linked markers to the trait of interest. Multiple methods have been used to identify markers that are closely linked to a gene of interest including marker analysis of near-isogenic-lines and bulk segregant analysis. In an $F_2$ population segregating for a known major gene of interest, bulk segregant analysis is the most effective way to identify a closely-linked marker.

Numerous studies have been conducted, using BSA and AFLP to map a gene of interest. Jong et al. (1997) mapped the *Nb* gene that confers resistance in potato *(Solanum tuberosum*) to potato virus X. In this study they screened parents and bulks with 96 AFLP primer combinations to identify eight putatively linked markers. Of these, three AFLP bands were successfully converted to PCR-based cleaved amplified polymorphic sequence (CAPS) markers. These markers were used in the mapping of the Nb locus to potato chromosome V.

In a similar study, Bendahmane et al.(1997) used AFLPs to develop a high resolution map around the *Rx*1 locus for extreme resistance in potato to potato virus X. The *Rx*1 locus had been previously mapped to potato linkage group XII. In this study they screened a total of 728 primer pairs using BSA to identify 57 potentially linked markers. Of these they converted three tightly linked AFLPs to CAPS

markers in order to determine their chromosomal location in a large mapping population. The three markers surrounded the *Rx*1 locus and all three mapped within 0.5 cM of the gene. These markers should serve as valuable tools for the physical capture of the *Rx*1 locus.

In soybean genetic mapping, AFLP has proven to be a valuable marker strategy. Maughan et al.(1996) performed an AFLP study, evaluating 15 AFLP primer pairs for polymorphism in 12 *Glycine max* and 11 *Glycine soja* accessions. Of the 759 fragments detected, 274 (36%) were polymorphic between at least two accessions. While this level of polymorphism can not be expected to be seen in an intraspecific cross of closely related cultivars, the data suggest that AFLP is the most powerful and efficient marker strategy available for detecting polymorphism among lines.

Keim et al. (1997) have recently published an AFLP-based linkage map in soybean. The map consists of 28 linkage groups covering 3441 cM. Six-hundred-fifty AFLP markers were mapped. The map contains a marker density of 1 per 4 cM, but there are large gaps between markers in certain regions of the genome. This suggests that AFLP markers may not be as ubiquitous as once postulated.

*Rsv*4 is an important resistance gene both because it confers broad resistance to SMV and also because its mode of

action, though not entirely understood, appears to be distinct from the hypersensitive response type of disease resistance. Because of its unique resistance nature there is interest in incorporating this along with other resistance loci such as *Rsv*1 and *Rsv*3 into a single cultivar, which would then, presumably possess multiple lines of defense against SMV infection. The ability to pyramid resistance genes into a single cultivar is greatly expedited by the use of closely linked molecular markers. Since genes such as *Rsv*4 and *Rsv*1 mask one another's presence, selecting lines that contain both genes is not possible by simple phenotypic methods. By using genetic marker-assisted selection, gene pyramiding becomes a viable breeding alternative for introducing multiple disease resistance.

In addition, there is tremendous interest in understanding how resistance genes are involved in defense response against invading pathogens. Numerous disease resistance genes have been cloned in the last five years, and a large number of these genes share common nucleotide binding site (NBS) and leucine rich repeat (LRR) motifs. All of these NBS-LRR genes confer resistance by means of a hypersensitive response where-by localized cell death is triggered in response to pathogen ingress, in order to deter

the spread of the pathogen. *Rsv*4 appears phenotypically to be distinct from this large class of resistance genes because it produces a symptomless response in the infected plant (i.e. no observable necrosis). Because of the unique nature of this gene, it would be of interest to clone it in order to better understand the mechanism controlling resistance. The first step toward cloning any gene of interest such as *Rsv*4 is to map its location within the genome.

In this study the objective is to use AFLP and BSA to identify a marker linked to the previously unmapped *Rsv*4 locus. This marker will be mapped in a reference population to identify the soybean linkage group containing *Rsv*4. Finally, RFLP and SSR markers will be identified and mapped on the *Rsv*4 linkage group to develop a linkage map surrounding this important resistance locus.

**Materials and Methods**

Soybean cultivar Lee68 (*rsv*4) was crossed with the breeding line D26-1-1(*Rsv*4) to produce an $F_2$ population of 281 individuals (DL population). Disease data for strains G1 and G7 have been previously collected on 15 to 20 plants of all 281 $F_{2:3}$ lines as described (Ma, 1996). This population was used for the fine mapping of the *Rsv*4 locus.

A second mapping population of 149 individuals of the cross V71-370 X PI407162 (VP population (Maughan et al., 1996)) that contains over 350 mapped loci, was used as a reference population to assign *Rsv*4 to a previously defined linkage group.

In the fall of 1995, 15 to 20 seeds of each $F_{2:3}$ line in addition to parental lines were planted in the greenhouse. Young trifoliate leaves of all healthy germinating plants were collected 2 to 3 weeks after planting. DNA was isolated by freeze-drying followed by CTAB extraction as described (Saghai Maroof et al., 1984).

DNA from twelve lines which are homozygous resistant and twelve lines which are homozygous susceptible was pooled to form bulk R and bulk S samples. Dual bulk R and bulk S DNA samples were made representing two mutually exclusive R and S subsets of the population. This was done in order to more precisely confirm the detection of a linked locus.

AFLP analysis was conducted according to Vos et al. (1995) and Maughan et al. (1996). Briefly bulk and parental DNA was double digested with restriction enzymes *Eco*RI and *Mse*I. Specific *Mse* and *Eco* adaptors were ligated to the restriction fragment ends. A subset of the fragments were amplified using adaptor-specific primers containing a single selective base at the 3' end (+1 reaction). A second round

of PCR was conducted in which $^{32}P$ end-labeled *Eco* primers and unlabeled *Mse* primers were used (+3 reaction). Each selective primer contained three selective nucleotides at the 3' end. Second round PCR reactions were run on 7 M urea, 6% polyacrylamide gels for three hours at 45 watts. The gel was then transferred to 3MM blot paper, covered with plastic wrap and exposed to X-ray film for three hours. Alternatively, selective amplification was conducted using a fluorescently-labeled *Eco* primer. Amplified products were run on a 6 M urea, 4.5% polyacrylamide gel for five hours at 1680 V and visualized using the ABI 377 gene scan system (Applied Biosystems; Foster City, CA). Primers used for AFLP screening are as listed in Table 1.

AFLP bands were converted to RFLP probes by excision of the polymorphic band from the polyacrylamide gel. Bands cut from the gel were eluted in 100 uL of water incubated in a boiling bath for 15 min. (Upender et al., 1995). After elution, a small aliquot was used for PCR amplification of the excised DNA fragment. This PCR product was then cloned into the pCNTR shuttle vector using the General Contractor cloning kit from 5prime-3prime (Boulder, CO). Subsequent AFLP reactions were conducted on plasmid DNA, as described above, of putative positive clones. In this case, 50 ng of plasmid template were used instead of genomic template. The

fragments were then run alongside parental DNA on a 7 M urea, 6% polyacrylamide gel for three hours at 45 watts to tentatively confirm that the proper band was cloned.

**Table 1.** Number of AFLP bands polymorphic between the lines D26-1-1 and Lee68. Polymorphisms were detected by testing 101 *Eco* +3/*Mse* +3 primer combinations using AFLP on parents and bulks of the population D26-1-1 X Lee68. Boxes marked in gray represent primer combinations that were not tested.

| ***Eco*** / ***Mse*** | **AAA** | **AAC** | **AAG** | **AAT** | **ACA** | **AGA** | **ACC** | **AGG** | **ACG** | **AGC** |
|---|---|---|---|---|---|---|---|---|---|---|
| **CAC** | 3 | 3 | 3 | 1 | 2 | 4 | 1 | 0 | 6 | 5 |
| **CAG** | 3 | 5 | 3 | 0 | 0 | 3 | 4 | 1 | 7 | 1 |
| **CAT** | 0 | 1 | 0 | | 1 | 0 | 0 | 1 | 2 | 2 |
| **CGA** | 0 | 2 | 0 | 0 | 0 | 0 | 1 | 1 | 0 | 1 |
| **CGT** | 2 | 2 | 3 | 4 | 1 | 1 | 1 | 0 | 1 | 0 |
| **CCT** | 0 | 0 | 1 | 0 | 4 | 1 | 1 | 1 | 2 | 2 |
| **CTG** | 2 | 1 | 2 | 1 | 0 | 0 | 2 | 1 | 1 | 2 |
| **CTT** | 0 | 2 | 1 | 0 | 1 | 1 | 1 | 0 | | |
| **AAG** | 1 | 1 | 2 | 1 | | | | | | |
| **AAT** | 1 | 0 | 1 | 1 | | | | | | |
| **AGA** | 3 | 1 | 0 | 0 | | | | | | |
| **AGG** | 1 | 0 | 4 | 3 | | | | | | |
| **ACG** | 0 | 0 | 1 | 1 | | | | | | |
| **ATA** | 1 | 1 | 3 | 2 | | | | | | |

The cloned PCR product was used as a probe for RFLP analysis essentially as previously described (Yu et al., 1996). Briefly, 8 µg of parental DNA was digested with enzymes *Dra*I, *Eco*RI, *Eco*RV, *Hin*dIII, *Xba*I, and *Taq*I, according to manufacturers protocols. Digested DNA was then separated on a 1% agarose gel at 70 to 90 mAmps for 14 to 16 h. The DNA was transferred to Hybond nylon membrane (Amersham, Piscataway, NJ) by Southern blotting with 0.4 N NaOH buffer. Screening blots were hybridized with $\alpha^{32}$PdCTP, random primer labeled probe. Hybridizing bands were visualized by autoradiography on Kodak (New Haven, CT) Xomat film.

In addition SSR analysis was conducted essentially as Cregan (1998). Briefly 50 ng of parental and $F_2$ DNA was used as template in a 20 uL reaction containing 1X reaction buffer (10 mM Tris-HCL, 50 mM KCL, pH 8.3), 2.5 mM MgCl2, 2 uM of each primer, 50 uM each of dATP, dGTP, dTTP, 1 uM of dCTP, and 1 uM of $\alpha$32P-dCTP, and 1.0 unit of Taq polymerase. Thirty cycles of a standard PCR reaction were run with denaturation at 94°C for 30 sec., primer annealing at 50°C for 30 sec., and primer extension at 72°C for 60 sec. Primers for SSR analysis were obtained from Research Genetics Inc. (Huntsville, AL) or custom made by Gibco-BRL Life Technologies (Rockville, MD).

Genetic mapping was conducted using the MAPMAKER computer program (Lander et al., 1987). Analysis was conducted at a threshold LOD of 3.0.

**Results**

AFLP analysis of the DL (D26-1-1 X Lee68) population showed a very low level of polymorphism. A total of 101 primer combinations were screened with DNA from parents and bulks (Table 1). A mean of 50 clearly distinguishable AFLP bands were detected per combination. Of these, only 141 total polymorphisms were detected between the parents among all primer combinations tested. This is equivalent to approximately 1.4 polymorphisms per AFLP reaction. This equates roughly to a genome coverage of approximately one polymorphic marker every 22 cM based on an average predicted soybean genome size of roughly 3100 cM (Cregan, 1998; Keim, 1997). Of these 141 polymorphisms, two were determined to be linked to the Rsv4 gene by bulk segregant analysis. These linked polymorphisms, R4-1 and R4-2, were detected with the primer combinations *Eco*+AAA/*Mse*+AAT and *Eco*+AAA/*Mse*+AGG, respectively.

R4-1 is an easily distinguishable dominant band of ~250 bp, that is absent in Lee68 and bulk S and present in D26 and bulk R (Fig. 1). R4-1 was excised by elution and cloned

into the pCNTR shuttle vector using the General Contractor 5prime-3prime cloning kit. When used as a probe, the R4-1 clone hybridized to a single polymorphic band on a Southern blot containing parental and bulk DNA digested with enzymes *Eco*RI and *Dra*I. Both the D26 parent and resistant bulk were of one allele while Lee68 and the susceptible bulk were of a

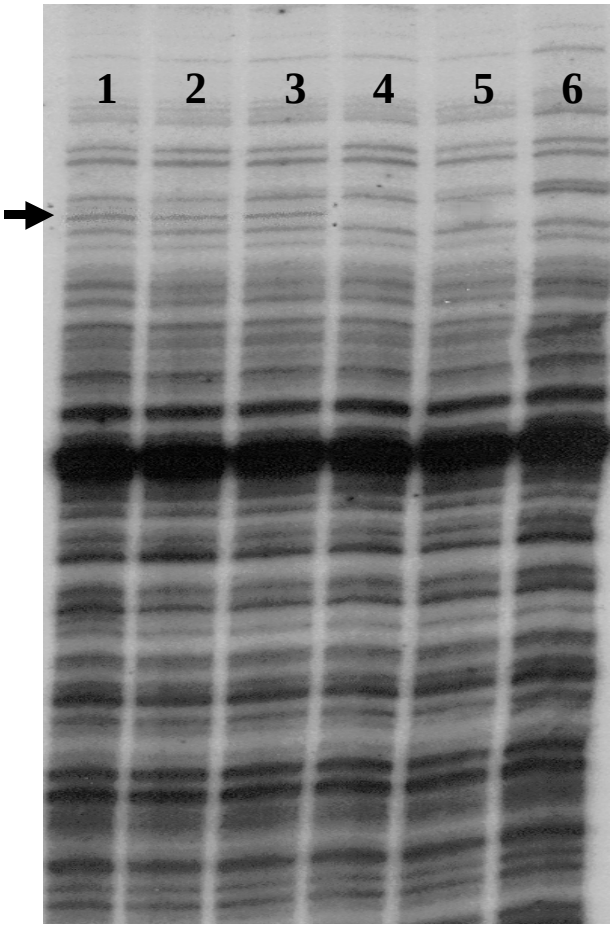



**Figure 1.** AFLP film of DL parents and bulks using primer combination *Eco*+AAA/*Mse*+AAT. The arrow indicates the linked polymorphic band designated R4-1, that is present in lanes 1,2 and 3. Lanes are: 1=D26-1-1 (*Rsv*4), 2&3=resistant bulks, 4&5=susceptible bulks, 6=Lee68 (*rsv*4)).

second allele. The clone was mapped in the DL population using the enzyme *Eco*RI (Fig. 2) and was found to be located 4.8 cM from the Rsv4 locus. Similarly the probe was mapped in the VP population with enzyme *Dra*I which showed an equivalent sized single-copy polymorphic band. Linkage analysis showed that the R4-1 probe maps to the soybean linkage group D1b, 25.8 cM from the RFLP marker A605. Linkage of A605 to Rsv4 was then confirmed by mapping the A605 probe in the DL population in order to verify that Rsv4 is located on the D1b linkage group. A605 was shown to be linked to Rsv4 at a distance of 26.3 cM (Fig. 3).

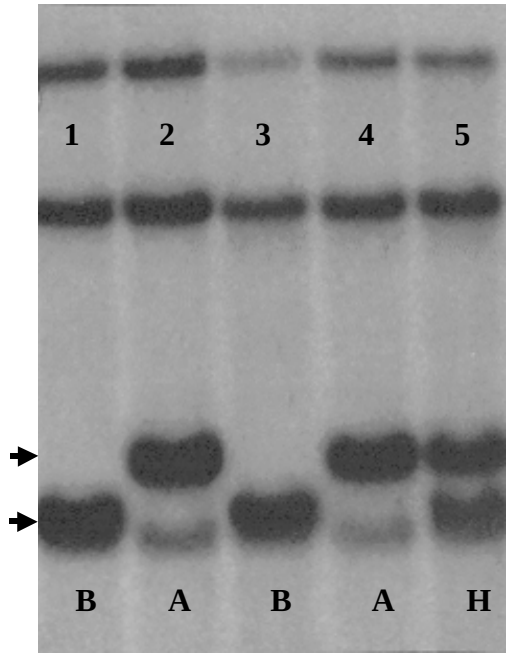



**Figure 2**. Southern blot of parents D26-1-1 and Lee 68 as well as 3 $F_2$s digested with restriction enzyme *Eco*RI showing hybridization with the AFLP converted RFLP probe R4-1. Arrows indicate *Rsv*4-linked polymorphic band. Lanes are: 1=Lee68 (Susc.), 2=D26 (Res.), 3,4,&5=$F_2$s (Homozygous Susc., Homozygous Res., and Heterozygous respectively). A&B = homozygous individuals, H=heterozygous individual.

The second AFLP marker R4-2 is a codominant AFLP product of approximately 120 bp. This particular band was only weakly amplified that hampered efforts to

convert it to an RFLP clone. It was instead mapped as an AFLP marker in a 96-sample subset of the DL population. Using this data set, it was determined that R4-2 is located

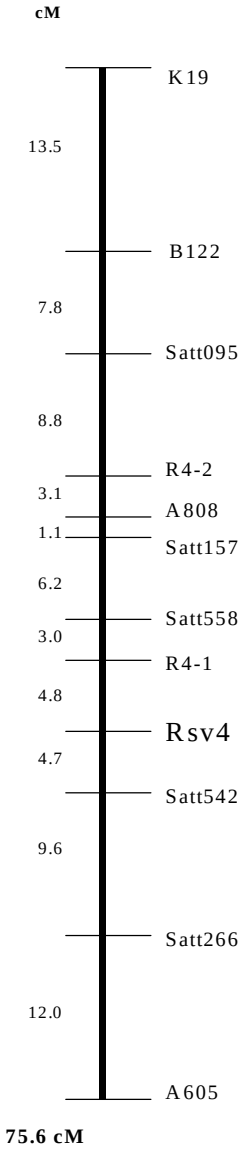



**Figure 3**. Map of linkage group D1b surrounding the SMV resistance locus, *Rsv*4. Map data is calculated based on a 255 individual $F_2$ population of the cross D26-1-1 X Lee68.

18 cM from Rsv4 on the same side as R4-1. Because of the distant linkage relationship, no further attempts were made to clone this particular fragment.

Based on the D1b linkage group information, several published RFLP and SSR markers were identified as likely candidates to be linked to Rsv4 (Cregan et al., 1998). Markers were screened and polymorphic loci including RFLP marker pK19 and SSRs Satt095, Satt157, Satt558, Satt542, and Satt266 were mapped in the DL population. Rsv4 was mapped to a region of D1b which is flanked by SSR markers Satt542 at 4.7 cM and Satt558 at 7.8 cM.

**Discussion**

Mapping of the Rsv4 locus was significantly expedited by the use of AFLP and bulk segregant analysis. This was particularly useful because of the low level of polymorphism observed in this particular population. Earlier studies have estimated the level of polymorphism detected by AFLP. Maughan et al.(1996) observed 127 polymorphisms out of 759 bands detected in at least one of twelve *G. max* lines evaluated. In this particular study, we obtained 141 clear polymorphisms from approximately 5050 amplified fragments. This works out to an extremely low level of polymorphism of

2.7%. This low level of polymorphism is likely due to the fact that both D26 and Lee68 are closely related to cv. Essex, such that their genomes are likely to be very homologous.

In this study, we were able to show that Rsv4 maps to the D1b chromosome. Previous work has indicated that numerous disease resistance loci are clustered in various regions of the soybean genome. Rsv1, for instance, is located on the F linkage group, in a small genomic region where genes for resistance to the pathogens *Phytophthora soja* (Rps3)(Diers et al., 1992), *Pseudomonas syringae* pv. *glycinea* (Rpg1) (Ashfield et al., 1998), and peanut mottle virus (Rpv)(Roane et al., 1983) have been identified. It is of interest to note that Rsv4, by contrast, is located in a portion of the genome where no other disease resistance locus has been identified.

The mapping of the *Rsv*4 locus has resulted in the identification of closely linked markers that will be useful in the marker-assisted selection of lines containing this gene. In addition we have laid the groundwork leading toward the eventual cloning of this important gene. Future efforts will involve identifying more closely linked markers in order that physical mapping of the *Rsv*4 region can be conducted.

**Literature Cited**

Ashfield, T., J. R. Danzer, D. Held, K. Clayton, P. Keim, M. A. Saghai Maroof, D. M. Webb, and R. W. Innes. 1998. Rpg1, a soybean gene effective against races of bacterial blight, maps to a cluster of previously identified disease resistance genes. Theoretical and Applied Genetics 96:1013-1021.

Bendahmane, A., K. Kanyuka, and D. C. Baulcombe. 1997. High-resolution genetical and physical mapping of the Rx gene for extreme resistance to potato virus X in tetraploid potato. Theoretical and Applied Genetics 95:153-162.

Buzzell, R. I. and J. C. Tu. 1989. Inheritance of a soybean stem-tip necrosis reaction to soybean mosaic virus. Journal of Heredity. 80:400-401.

Chen, P., G. R. Buss, C. W. Roane, and S. A. Tolin. 1994. Inheritance in soybean of resistant and necrotic reactions to soybean mosaic virus strains. Crop Science 34:414-422.

Cregan, P. B., T. Jarvik, A. L. Bush, R. C. Shoemaker, K. G. Lark, A. L. Kahler, T. T. Van Toai, D. G. Lohnes, J. Chung, and J. E. Specht. 1998. An integrated genetic linkage map of soybean. Crop Science *in press*.

Diers, B. W., L. Mansur, J. Imsande, and R. C. Shoemaker. 1992. Mapping *phytophthora* resistance loci in soybean with restriction fragment length polymorphism markers. Crop Science 32:377-383.

Irwin, M. E., and R. M. Goodman. 1981. Ecology and control of soybean mosaic virus. pp. 181-220 in: Maramorosch, K. and K. Harris (eds.) Plant diseases and vectors: ecology and epidemiology. Academic Press, New York.

Jong, W. D., A. Forsyth, D. Leister, C. Gebhardt, and D. C. Baulcombe. 1997. A potato hypersensitive resistance gene against potato virus X maps to a resistance gene cluster on chromosome 5. Theoretical and Applied Genetics 95:246-252.

Keim, P., J. M. Schupp, S. E. Travis, K. Clayton, and D. M. Webb. 1997. A high-density soybean genetic map based upon AFLP markers. Crop Science 37:537-543.

Kihl, R. A. S., and E. E. Hartwig. 1979. Inheritance of reaction to soybean mosaic virus in soybeans. Crop Science 19:372-375.

Lander, E. S., P. Green, J. Abrahamson, A. Barlow, J. M. Daly, S. E. Lincoln, and L. Newberg. 1987. Mapmaker: an interactive computer package for constructing primary genetic linkage maps of experimental and natural populations. Genomics 1:174-181.

Ma, G. 1996. Genetic analysis of soybean reactions to soybean mosaic virus. PhD dissertation. Virginia Tech.

Ma, G., P. Chen, G. R. Buss, and S. A. Tolin. 1995. Genetic characteristics of two genes for resistance to soybean mosaic virus in PI486355 soybean. Theoretical and Applied Genetics 91:907-914.

Maughan, P. J., M. A. Saghai Maroof, and G. R. Buss. 1996. Molecular-marker analysis of seed-weight: genomic locations, gene action, and evidence of orthologous evolution among three legume species. Theoretical and Applied Genetics 93:574-579.

Maughan, P. J., M. A. Saghai Maroof, G. R. Buss, and G. M. Huestis. 1996. Amplified fragment length polymorphism (AFLP) in soybean: species diversity, inheritance, and near-isogenic line analysis. Theoretical and Applied Genetics 93:392-401.

Michelmore, R.W., I. Paran, and R. V. Kesseli. 1991. Identification of markers linked to disease-resistance genes by bulked segregant analysis: a rapid method to detect markers in specific genomic regions by using segregating populations. Proceedings of the National Academy of Sciences, USA 88:9828-9832.

Ren, Q. T. W. Pfeiffer, and S. A. Ghabrial. 1997. Soybean mosaic virus incidence level and infection time: Interaction effects on soybean. Crop Science 37:1706-1711.

Roane, C. W., S. A. Tolin, and G. R. Buss. 1983. Inheritance of reaction to two viruses in the soybean cross ‘York’ x ‘Lee 68'. Journal of Heredity 74:289-291.

Saghai Maroof, M. A., K. M. Soliman, R. A. Jorgensen, and R. W. Allard. 1984. Ribosomal DNA spacer length polymorphism in barley: Mendelian inheritance, chromosomal location, and

population dynamics. Proceedings of the National Academy of Sciences, USA 81:8014-8018.

Thottapilly, G. and H. W. Rossel. 1987. Viruses affecting soybean *in* Singh, S. R., K. O. Rachie, and K. E. Dashiell (eds.). Soybeans for the tropics: research, production, utilization. John Wiley and Sons Ltd., Chichester, U.K. pp.53-68.

Upender, M., L. Raj, and M. Weir. 1995. Rapid method for the elution and analysis of PCR products separated on high resolution acrylamide gels. Biotechniques 18:33-34.

Vos, P., R. Hogers, M. Bleeker, M. Reijans, T. van de Lee, M. Hornes, A. Frijters, J. Pot, J. Peleman, M. Kuiper, and M. Zabeau. 1995. AFLP: a new technique for DNA fingerprinting. Nucleic Acids Research 23:4407-4414.

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

# CHAPTER III

# HOMOLOGY-BASED CLONING OF A CANDIDATE DISEASE RESISTANCE GENE IN SOYBEAN

**Abstract**

The cloning of several plant genes directly involved in triggering a resistance response has shown that numerous resistance genes have similar conserved amino acid sequence. This research demonstrates the potential to clone disease resistance genes from related crops using a homology-based approach. Using NBS5, a soybean resistance-gene-like clone that maps to a cluster of resistance genes on linkage group F, we were able to identify two putative genes, one of which appears to be functional based on nucleotide sequence information. The putative functional gene contains a 3,697 nucleotide open reading frame (ORF) sequence which when translated reveals an N-terminal region with a high level of similarity to the NBS/LRR class of cloned disease resistance genes. The C-terminal region is more divergent as seen in other cloned genes, but it does contain a direct imperfect repeat and leucine rich repeat (LRR) region that is characteristic of cloned disease resistance genes. A second cloned gene that is homoeologous to the F-linked gene is located on the E-linkage group. It contains 1,565 nucleotides of ORF in the N-terminal domain. The C-terminal region of this gene however contains numerous stop codons and apparent frame shifts, suggesting that its functionality as a disease resistance gene is questionable. These genes,

L20a and L33, are shown to be closely related to one another and to the members of the NBS/LRR class of disease resistance genes.

**Introduction**

The ultimate goal of mapping any disease resistance gene is to eventually identify the nucleotide sequence which codes for the resistance gene product (i.e., clone the gene). Ground breaking work employing both map-based and transposon tagging strategies has resulted in the cloning of a number of resistance genes (for a review see Baker et al., 1997). Several of these genes, including *N* from tobacco (*Nicotiana tabacum*) (Whitham et al., 1994), *L*6 from flax (*Linum usitatissimum*) (Lawrence et al., 1995), and *Rps2* from *Arabidopsis* (Bent et al., 1994), contain short amino acid sequences that are highly conserved.

Yu et al. (1996) designed degenerate primers from conserved sequences corresponding to two coding regions from the disease resistance genes, *N* and *Rps2*. They were able to amplify corresponding sequences of about 340 bp from soybean containing similar conserved sequences. These sequences contained conserved nucleotide binding site (NBS) motifs that consist of P-loop, kinase-2, and putative kinase-3a conserved motifs. At least eleven different resistance gene like (RGL) classes were identified by Yu et al. (1996). In a very similar study conducted by Kanazin et al. (1996), additional soybean RGL sequences were characterized. Several RGL classes have been mapped to regions of the

soybean genome where resistance genes have been identified. For example, two NBS classes, b and j, map to a resistance gene cluster on linkage group F that includes the *Rsv1*, *Rps3*, *Rpv*, and *Rpg1* loci (Yu et al., 1996).

Similar homology-based studies have been published by groups working with such diverse species as potato (*Solanum tuberosum*) (Leister et al., 1996), tomato (*Lycopersicon esculentum*) (Ohmori et al., 1998), *Arabidopsis* (Aarts et al., 1998; Speulman et al., 1998), sunflower (*Helianthus annuus*) (Gentzbittel et al., 1998) lettuce (*Lactuca sativa*) (Shen et al., 1998) and the grain crops, rice (*Oryza sativa*), wheat (*Triticum aestivum*), barley (*Hordeum vulgare*), and maize (*Zea mays*) (Collins et al., 1998; Leister et al., 1998; Seah et al., 1998). In all of these studies the researchers were able to identify RGL sequences that mapped very close to or cosegregated with resistance genes of interest. Sequence databases show similar RGL sequences reported in oat (*Avena sativa*) and pine *(Pinus radiata*)(Cheng and Armstrong, unpublished; McCallum et al., unpublished). In all, no less than 150 RGL sequences have been reported from various sources. Comparison of these sequences along with the corresponding conserved sequences from previously cloned resistance genes, shows that very different plant species have very similar types of RGL

sequences. Potato RGL clone, STU60080 for example, shares 52.4% structural similarity with the lettuce RGL clone, AF017754 and 52.9% similarity with soybean RGL clone, NBS17. In addition, there is evidence that within a species RGL sequences can be highly divergent. For instance *Arabidopsis* RGL sequence AF039377 is only 10.2% similar to another *Arabidopsis* RGL clone, AF039385, and is actually more similar to a group of potato RGL sequences. In general however, it is not surprising to notice that the stuctural tree of RGL sequences shows that most cluster into species specific groups.

Following analysis of the various RGL sequences cloned from soybean it was determined that the class b clone, NBS5, could represent sequence from an important gene involved in disease resistance. NBS5 is 36.4% similar to the NBS region of the *N* gene based on structural comparison. Its close proximity to a cluster of resistance genes on the soybean F linkage group, including *Rsv*1, *Rpv*, *Rpg*1, and *Rps*3, makes it a potentially important RGL sequence that might represent a functional resistance gene. In addition, the fact that it is essentially single copy upon Southern hybridization suggests that this particular clone might prove more easily cloned than some other high-copy, multi-gene family, NBS clones.

The objectives of this research are to identify a full length soybean gene which corresponds to the 336 bp NBS5 sequence, and to then characterize this gene as a candidate member of the NBS/LRR class of disease resistance genes.

## Materials and Methods

### Library Screening

Two lambda cDNA libraries and one lambda genomic library were screened in this study. The first cDNA library was constructed by Clontech (Palo Alto, CA) using the lambda tripleEx™ vector. It was produced from mRNA of 10-day-old leaf tissue of greenhouse grown plants of line L81-4420 which contains the Rsv1 gene derived from PI96983. The cDNA for producing this library were obtained by reverse transcription using both random and poly-T primers. A second library, obtained from Stratagene (La Jolla, CA), was constructed in the lambda Uni-zap XR™ vector. This library was produced from mRNA obtained from 12-day-old epicotyl tissue of greenhouse grown plants of the cultivar Williams 82. In this library, only poly-T primer was used for cDNA construction. A single lambda genomic library of the cultivar Williams 82 was screened. This library was constructed by Stratagene using the lambda FixII™ vector.

Libraries were plated out onto 150 mm plates according to manufacturer's protocols. Positive plaques were identified by transfer to Nylon membranes, followed by hybridization with $^{32}$P-labeled probe, NBS5 (and later the partial cDNA L2-5i). Briefly, Magnalift (MSI, Westboro, MA) nylon membranes were positioned onto freshly grown plates for three to four minutes. They were then washed for five minutes in a denaturing buffer (1.5M NaCl, 0.5N NaOH) followed by a five minute wash in a neutraliziation buffer (1.5 M NaCl, 0.5 M trisHCl pH7.5). Finally, the blots were rinsed in 2X SSC before being dryed and baked at 90°C for 2 h. Lifts were then prehybridized for 20 hours in (1 M $NaH_2PO_4$, 1 M $Na_2HPO_4$, 1 M PB buffer, 100 Denhardt's solution, and 5 mg/mL salmon sperm DNA). Random-primer labeled probe was added to the buffer and the lifts were incubated for 20 h at 65°C. Blots were then washed twice for five minutes with cold low stringency buffer (2X SSC, 0.5% SDS), followed by a fifteen minute wash with a higher stringency buffer (0.5X SSC, 0.1% SDS) at 65°C, and a second five minute wash under the same conditions. Lifts were covered with plastic wrap and exposed to Kodak (New Haven, CT) Xomat X-ray film for 8 to 24 hours. Positive plaques were excised from primary plates and were screened a second time using 90 mm plates at low plaque density to isolate single positives.

Evaluating positive clones

Lambda cDNA positives were converted to plasmid according to manufacturer's protocol for further analysis. Lambda DNA from the FixII™ genomic library was extracted using the Promega Wizard™ lambda DNA extraction kit, or conversely by $ZnCl_2$ precipitation according to Ming-Tsan et al.(1998).

Positive cDNA clones were mapped in two soybean populations. The population V71-370 X PI407162 (VP) was used for identifying the map location of positive clones (Maughan et al., 1996). This population of 149 individuals contains over 350 mapped loci covering all 20 linkage groups (Saghai Maroof, unpublished). A second population from the cross PI96983 X Lee68 (PL) was used for fine mapping of the Rsv1 region. This population contains 243 $F_2$ individuals segregating for the *Rsv*1 locus, and has at least 40 markers placed in the *Rsv*1 region of the F-linkage group.

Parental diagnostic and $F_2$ Southern blots were made as previously described (Yu et al., 1996). Briefly, 8 µg of parental and/or $F_2$ DNA were digested with enzymes *Dra*I, *Eco*RI, *Eco*RV, *Hind*III, *Xba*I, and *Taq*I according to manufacturers protocols (Gibco-BRL, Rockville, MD). Digested DNA was then separated on a 1% agarose gel at 70 to 90 mAmps for 14 to 16 hours. The DNA was transferred to

Hybond™ nylon membrane (Amersham, Piscataway, NJ) by Southern blotting with 0.4 M NaOH buffer. Screening blots were hybridized with $\alpha^{32}$PdCTP, random-primer labeled probe (Ambion, Austin, TX). Hybridizing bands were visualized by autoradiography on Kodak Xomat™ film. Bands were mapped based on genetic distances calculated by MapMaker 4.0 at LOD 3.0 (Lander et al., 1987).

Positive cDNA clones were sequenced either at the University of Georgia Molecular Genetics facility using an ABI373, or in-house using an ABI377 DNA sequencer (Applied Biosystems, Foster City, CA). Plasmid template was prepared by a standard alkaline-lysis procedure followed by purification using QiaexII (Qiagen, Valencia, CA). Sequencing was performed based on Perkin Elmer recommendations using dye-labeled terminators in a cycle sequencing reaction. Sequence assembly and analysis was conducted using Lasergene software from DNAstar (Madison, WI).

**Results**

Identifying a partial cDNA corresponding to NBS5

The initial screening with the NBS5 clone was conducted in the Fall of 1996. A Clontech cDNA library, custom made from the line L81-4420, was screened with the 336 bp NBS5

insert amplified with specific internal primers (tnk005R&F) using NBS5 plasmid as template. Originally 23 150 mm plates, with an average density of ~1500 plaques/plate, were screened by hybridization of lifts in triplicate. In this original screen a single positive clone, L2-5i, was isolated. This clone was evaluated during the Fall of 1996. After conversion of this clone to plasmid, digestion with *Eco*RI excised a single 2 kb band. Southern hybridization demonstrated that this clone hybridized to the same restriction fragment (*Eco*RI in VP) as the smaller NBS5 clone and correspondingly mapped to the same locus.

Sequence analysis of the L2-5i insert revealed a clone of 1,947 nucleotides in length. The clone contained 1,569 base pairs (bp) of open reading frame (ORF) sequence preceded by 52 bp 5' of the start site, interrupted by a single 123 bp putative intron, and followed by 203 bp of a second putative intron. The first exon contained the entire NBS5 sequence. Additionally sequences outside of the NBS region showed conservation with motifs identified from other previously cloned disease resistance genes. Sequence comparison was made with a partially sequenced soybean genomic clone, B23, previously identified by Yu et al. (1996), and it was confirmed that this clone corresponded very closely to sequence data of the genomic clone. After

analysis it was hypothesized that L2-5i represented a partial length genomic contaminate sequence (i.e., genomic DNA contaminating the cDNA library) corresponding to a putative resistance gene.

Additional cDNA library screening with L2-5i

In order to obtain a full length cDNA we used the L2-5i clone as a probe to screen the same library. Plaque density was increased to ~17500 plaques per 150 mm plate in order to increase the number of positives identified. This screen was conducted in the late Fall of 1996. In the primary screen, 32 plaques were identified as potential positives. After secondary screening, 22 clones were converted to plasmid for further analysis. After insert amplification of at least 5 clones from each putative positive convertant, no cDNA was identified that was larger than the original 2 kb clone, L2-5i. Four positives were confirmed by hybridization followed by full or partial sequencing. All of these were determined to contain portions of the same or essentially the same sequence as L2-5i. The positives analyzed were L75-5L, L76-5L, L78D-5L, and L85-5L. The complete sequencing of the L75-5L clone confirmed that L2-5i did indeed contain 123 bp of intron sequence. This was confirmed by the observation that both L75-5L and L85-5L contained continuous ORF sequence and lacked the 123 bp of

sequence found in L2-5i (all NBS5 cDNAs characterized at the sequence level in this project are summarized in Fig. 1). Since only 5' sequence of the various cDNAs had been identified in the initial two cDNA library screens and in the genomic clone previously reported, it became apparent that our focus should shift to obtaining sequence corresponding to the 3' end of the NBS5 gene sequence.

Screening of a cDNA library from cv. Williams 82

It was reasoned that since obtaining the 3' sequence was our major goal, we should screen a cDNA library made

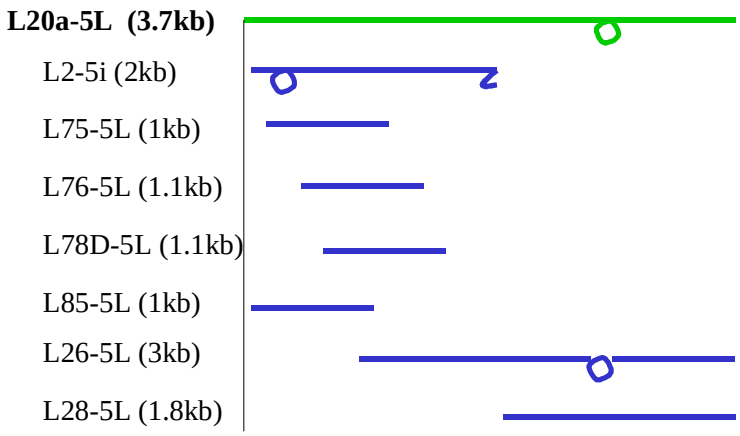



**Figure 1.** All F-linked cDNA clones sequenced showing which portion of the partial cDNAs is present as compared to the full-length clone L20a-5L. Clones were isolated from cDNA libraries of Williams 82 (L20a-5L, L26-5L, L28-5L) and L81-4420 (all others). Circles represent sequence present in some but not all clones.

only from poly-T primer. The lambda triplEx library was constructed using both poly-T and random primers. By ensuring that the poly-A tail was present we could be assured of obtaining a clone containing the 3' end of the NBS5 gene sequence. A soybean cDNA library of cv. Williams 82 was obtained from Stratagene in the winter of 1996 and screened in January of 1997. The library was screened with L2-5i using a plaque density of ~15000/plate. Forty-two plates were screened with duplicate lifts. A total of seven putative positives were excised for secondary analysis. Of these, six were continued for plasmid conversion and further analysis.

Identification of two types of cDNA clone mapping to E and F homoeologous chromosomal regions

The six positives identified were L9-5L, L10-5L, L21-5L, L26-5L, L28-5L, and L29-5L. Hybridization indicated that all the clones were similar, but with some very distinct differences. Clones L26 and L28 hybridized strongly to the same polymorphic restriction fragment (*Eco*RI in VP) previously mapped to linkage group F using NBS5 (Fig. 2). The other four clones hybridized weakly to this same restriction fragment. Conversely these four clones hybridized strongly to a second polymorphic fragment in which L26 and L28 hybridized very weakly. This band mapped

to a previously identified homoeologous region on linkage group E (Shoemaker et al., 1996). Based on these observations we deduced that these six NBS5 cDNA clones represent sequences from two unique loci. One gene sequence corresponding to NBS5 is coded for on chromosome F, while a second gene sequence, homoeologous to NBS5, is coded for on chromosome E.

Sequence analysis indicated that in fact the six sequenced clones did fall into two distinct sequence classes as previously described based on hybridization experiments. Not surprisingly, all of the six clones contained 3'

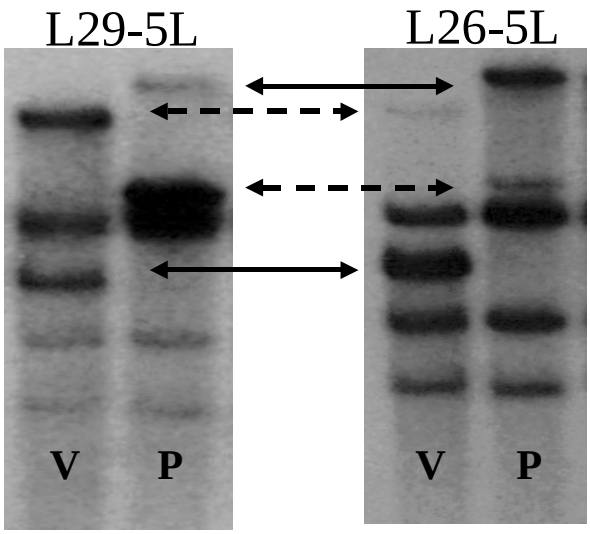



**Figure 2.** Southern blot of DNA from the mapping parents V71-370 (V) and PI407162 (P) digested with enzyme *Eco*RI. The same blot was hybridized with the putative E-linked cDNA clone, L29-5L and the putative F-linked cDNA clone, L26-5L. Dashed arrows indicate a polymorphic band mapping to linkage group E. Solid arrows indicate a polymorphic band mapping to linkage group F.

untranslated sequence indicating that we had fulfilled our goal of cloning the 3' region of the NBS5 gene. Disappointingly neither of the two F-linked clones, L26 or L28, was determined to be full length.

Evaluation of a full length chimeric contig of the F-linked clone

Both F-linked clones (L26-5L and L28-5L), from the Williams 82 library, did however contain overlap sequence with the original clone, L2-5i from the L81-4420 library. Consequently a working chimeric full-length contig was constructed using clones L28-5L and L2-5i. The ORF overlap length was approximately 269 bp. Within this overlap region there are ~12 apparent nucleotide discrepancies, presumably a product of the two cDNA clones being derived from different soybean lines. None of the nucleotide changes caused significant discrepancies in the deduced amino acid sequence, such as frame shifts or stop codons. Significant analysis and comparison was done between this chimeric contig and other resistance genes. These indicated upwards of 30% similarity between this cDNA sequence and tobacco resistance gene N.

Evaluation of the E-linked clones

In addition to the sequence analysis of the F-linked NBS5 clones, there was an effort to characterize the newly

identified E-linked clones, L9-5L, L10-5L, L21-5L, and L29-5L. L21-5L (4.5 kb) and L29-5L (2.5 kb) were completely sequenced. Because of its large size, L21-5L was postulated to be a full or near full length clone. In fact however, it was determined that a portion of the clone L21-5L was unrelated to NBS5. L21-5L resulted from the ligation of 3.2 kb of cDNA from the E-linked NBS5 sequence and 1.3 kb of unrelated soybean cDNA sequence. The unrelated 1.3 kb of cDNA sequence was highly homologous to a number of β-glucosidase genes that have been previously cloned from other crops (Oxtoby et al., 1991; Xue et al., 1992).

Another E-linked clone, L9, showed a similar chimeric clone structure. This clone contained sequence highly homologous to a proline-rich 14kDa putative protein cloned from kidney bean (*Phaseolus vulgaris*)(Choi et al., 1996). Similarly the L10 cDNA showed a chimeric structure, but the identity of the unrelated sequence could not be confirmed. Full or partial sequence information obtained on all four E-linked clones indicated that none of them were of full length.

Alternative approaches for identifying a full-length clone

While efforts to this point had resulted in the cloning of multiple cDNA clones, which when taken in concert, gave a clear picture of the sequence that makes up the NBS5 gene,

we still did not have a single clone containing the full gene. At this point, efforts were made to use all of the known sequence information to amplify either a full length clone from cDNA and genomic library template or just directly from genomic DNA template. Primers which had been designed during sequencing were used along with a pair of primers designed specifically from the ends of the cDNA clones. These primers were called FLNBS5F and FLNBS5R. While numerous bands were amplified from library and genomic DNA, these efforts were not successful in generating or detecting a full length clone.

Identification of a full-length F-linked and E-linked clone

A final effort was made in the Spring of 1997 to screen the Williams 82 cDNA library once again to obtain additional NBS5 related clones. A total of 54 150 mm plates with a plaque density of ~15000 per plate were screened with the clone L2-5i. Twelve putative positives were excised after the primary screening. Following secondary screening all twelve were converted to plasmid for further analysis. After PCR and/or restriction analysis and Southern hybridization, two clones were found to be identical to previously identified cDNAs. The clones L27-5L and L41-5L are repeats of L21-5L and L26-5L, respectively. In addition, two new large-insert positives, L20a-5L (3.7 kb)

and L33-5L (3.5 kb), were identified as potential full length clones. Southern hybridization indicated that L20a-5L was F-linked while L33-5L was E-linked. Both clones were sequenced in July of 1997. After sequence analysis it was confirmed that L20a-5L represented the full length cDNA corresponding to the original 336 bp NBS5 sequence.

Evaluation of full length, F-linked clone, L20a-5L

L20a-5L has an insert size of 3,697 nucleotides, including a 19 nucleotide poly-A tail. The putative start

| | |
|---|---|
| MSKAVSESTDIRVYDVFLSFRGEDTRRSFTGNLYNCLEKRGIHTFIGDYD | 50 |
| FESGEEIKASLSEAIEHSRVFVIVFSENYASSSWCLDGLVRILDFTEDNH | 100 |
| RPVIPVFFDVEPSHVRHQKGIYGEALAMHERRLNPESYKVMKWRNALRQA | 150 |
| ANLSGYAFKHGDGYEYKLIEKIVEDISNKIKISRPVVDRPVGLEYRMLEV | 200 |
| DWLLDATSLAGVHMIGIC**GIGGIGKTTLARAVYHSAAGHFDTSCFLGNVR** | 250 |
| **ENAMKHGLVHLQQTLLAEIFRENNIRLTSVEQGISLIKKMLPRKRLLLVL** | 300 |
| **DDVCELDDLRALVGSPDWFGPGSRVIITTR**DRHLLKAHGVDKVYEVEVLA | 350 |
| NGEALELLCWKAFRTDRVHPDFINKLNRAITFASGIPLALELIGSSLYGR | 400 |
| GIEEWESTLDQYEKNPPRDIHMALKISFDALGYLEKEVFLDIACFFNGFE | 450 |
| LAEIEHILGAHHGCCLKFHIGALVEKSLIMIDEHGRVQMHDLIQQMGREI | 500 |
| VRQESPEHPGKRSRLWSTEDIVHVLEDNTGTCKIQSIILDFSKSEKVVQW | 550 |
| DGMAFVKMISLRTLIIRKMFSKGPKNFQILKMLEWWGCPSKSLPSDFKPE | 600 |
| KLAILKLPYSGFMSLELPNFLHMRVLNFDRCEFLTRTPDLSGFPILKELF | 650 |
| FVFCENLVEIHDSVGFLDKLEIMNFEGCSKLETFPPIKLTSLESINLSHC | 700 |
| SSLVSFPEILGKMENITHLSLEYTAISKLPNSIRELVRLQSLELHNCGMV | 750 |
| QLPSSIVTLRELEVLSICQCEGLRFSKQDEDVKNKSLLMPSSYLKQVNLW | 800 |
| SCSISDEFIDTGLAWFANVKSLDLSANNFTILPSCIQECRLLRKLYLDYC | 850 |
| THLHEIRGIPPNLETLSAIRCTSLKDLDLAVPLESTKEGCCLRQLILDDC | 900 |
| ENLQEIRGIPPSIEFLSATNCRSLTASCRRMLLKQELHEAGNKRYSLPGT | 950 |
| RIPEWFEHCSRGQSISFWFRNKFPVISLCLAGLMHKHPFGLKPIVSINGN | 1000 |
| KMKTEFQRRWFYFEFPVLTDHILIFGERQIKFEDNVDEVVSENDWNHVVV | 1050 |
| SVDVDFKWNPTEPLVVRTGLHVIKPKSSVEDIRFIDPYKPTFL | 1093 |

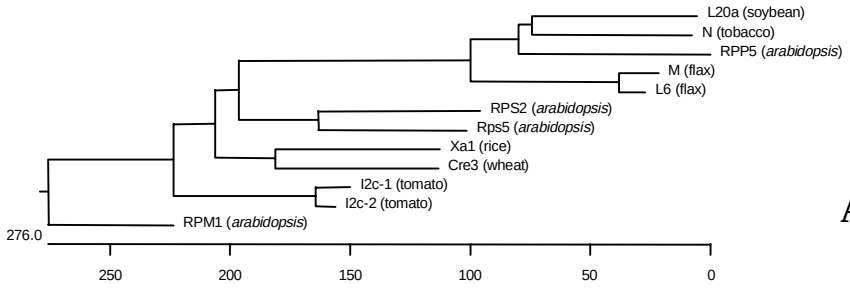
**Figure 3.** Deduced amino acid sequence of the open reading frame portion of cDNA clone L20a corresponding to the NBS5 sequence of soybean. The putative protein is 1092 amino acids in length. The nucleotide binding site portion of the sequence is in bold letters. The leucine rich repeat region is underlined.

site is located at position 73 and the ORF extends to a stop codon ending at position 3354. The total length of the NBS5 F-linked gene is 3,279 nucleotides which translates to a 1093 amino acid protein (Fig. 3). The clone contains 344 nucleotides of 3' untranslated sequence. Interestingly there is 546 nucleotides of sequence present in L20a-5L which is not present in the partial clone L28-5L. This suggests the possibility of two alternatively spliced mRNAs derived from this single gene sequence. Hybridization of L20a-5L with diagnostic southern blots showed the expected F-linked pattern, but revealed a new polymorphic band that had not been observed in any of the other partial cDNAs. This band maps to the E linkage group near RFLP marker pB212, in a location 45 cM from that of the homoeologous E-linked clones mapped in this study.

Sequence analysis indicates that L20a-5L shares numerous conserved, defined and undefined motifs with NBS-LRR type resistance genes. Based on clustal alignment analysis (Higgins and Sharp, 1989), L20a-5L is most similar to N, a virus resistance gene from tobacco, sharing 26% sequence similarity. It is least similar to Xa1, a bacterial resistance gene from rice, sharing only 12% similarity (Fig. 4).

Evaluation of full length, E-linked clone L33-5L

Sequencing of the second positive clone, L33-5L, indicated that it too represents a full length sequence which is homoeologous to L20a. L33 is 3,525 nucleotides in length including a 15 nucleotide poly-A tail. The start site for this clone is proposed to begin at position 82. The open reading frame for this clone only extends to position 1646. At this point, there is a single stop codon




**A**

Percent Similarity

| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | | 26.0 | 12.6 | 13.1 | 21.6 | 20.4 | 12.3 | 22.7 | 14.6 | 12.0 | 12.3 | 14.3 | 1 | L20a (soybean) |
| 2 | | | 11.9 | 12.2 | 20.8 | 20.7 | 13.0 | 25.0 | 12.8 | 11.7 | 13.8 | 14.2 | 2 | N (tobacco) |
| 3 | | | | 75.7 | 13.4 | 13.0 | 17.2 | 12.2 | 13.2 | 17.5 | 17.6 | 15.2 | 3 | I2c-1 (tomato) |
| 4 | | | | | 13.5 | 14.8 | 19.0 | 14.0 | 14.3 | 16.8 | 17.7 | 16.6 | 4 | I2c-2 (tomato) |
| 5 | | | | | | 70.0 | 14.3 | 16.9 | 13.4 | 12.6 | 12.5 | 11.6 | 5 | M (flax) |
| 6 | | | | | | | 13.9 | 17.2 | 14.0 | 12.5 | 12.0 | 10.9 | 6 | L6 (flax) |
| 7 | | | | | | | | 12.4 | 15.7 | 12.0 | 15.3 | 16.1 | 7 | RPM1 (*arabidopsis*) |
| 8 | | | | | | | | | 13.3 | 13.1 | 11.9 | 12.8 | 8 | RPP5 (*arabidopsis*) |
| 9 | | | | | | | | | | 13.0 | 14.2 | 27.0 | 9 | RPS2 (*arabidopsis*) |
| 10 | | | | | | | | | | | 18.1 | 13.4 | 10 | Xa1 (rice) |
| 11 | | | | | | | | | | | | 13.2 | 11 | Cre3 (wheat) |
| 12 | | | | | | | | | | | | | 12 | Rps5 (*arabidopsis*) |
| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | | |

**B**

**Figure 4.** A. Phylogenetic tree of L20a-5L and cloned NBS-LRR disease resistance genes showing relatedness based on deduced amino acid sequence of full-length genes. Comparisons were made using the Clustal alignment method (Saitou and Nei, 1987). B. Percent similarity table based on the same Clustal analysis.

followed by sequence that is highly characteristic of intron. This intron-like sequence extends for 250 bp. Following this there is severely disjointed "ORF" sequence which proceeds until a stop codon ending at position 3190. With the exception of the first 1,565 nucleotides of ORF sequence, this clone does not contain continuous translatable sequence. By comparing the sequence of L20a with L33, one could detect the continuation of translated sequence that shifted frames multiple times and even includes two point mutations to a stop codon. Notably there is ~250 bp of sequence present in L33 not present in L20a, and similarly ~300 bp present in L20a not detected in L33. It is clear however, that L33 does not code for a full length RGL protein product like L20a.

**Discussion**

In this study a gene was identified that corresponds to the 336 bp RGL sequence, NBS5 of soybean. After completely sequencing this gene as represented by the L20a-5L cDNA clone, sequence comparisons were done to determine the relatedness of this clone to other cloned disease resistance genes (Fig. 4). L20a-5L clearly belongs in the class of plant disease resistance genes that contain an NBS and LRR. The L20a-5L clone most closely resembles the N gene cloned

from tobacco that confers resistance to tobacco mosaic virus. Sequence similarity between these two genes is 26%. L20a-5L is in a subgroup of NBS-LRR resistance genes, which includes N, L6 (flax), M (flax), and Rpp5 (*Arabidopsis*). Similarities of these clones to L20a-5L range from 20.4% to 26%. A second subgroup of NBS-LRR genes which includes *Rps2* (*Arabidopsis*), *Rpm1* (*Arabidopsis*), I2 (tomato), Xa1 (rice), and the putative Cre3 gene (wheat) are similar to L20a-5L in the range of 12% to 14.6%. The N-terminal region of the translated L20a-5L sequence contains numerous conserved motifs, both defined and undefined, that are shared by the members of this group. These include the P-loop, kinase-2, and putative kinase-3a regions as defined by Lawrence et al. (1995), who also described other undefined regions of amino acid conservation. All the potentially significant conserved motifs shared between L20a-5L and other cloned resistance genes that are present in the N-terminal half of the protein are listed in Table 1.

In addition to N-terminal conservation in the putative protein product of L20a-5L, there are conserved characteristics in the COOH-terminal region which further suggest a disease resistance role for this gene product. Like all genes in this class, L20a-5L contains a region that resembles the common LRR found on all NBS-LRR resistance

genes in this portion of the protein product (Fig. 5). While no two LRRs in this superfamily of genes are identical, all share commonalities of structure. LRRs are hypothesized to play a role in protein-protein interactions, which suggests that this portion of the gene is involved in pathogen recognition (Baker et al., 1997). It is perhaps not surprising then that all of these resistance genes show slightly different LRR patterns, lending to their various specificities.

1990 L sg F pi L ke L f F vf C en L ve IH ds
V gf L dk L ei M n F eg C sk L et FP p
I -k L ts L es I n L sh C ss L vs FP ei
L gk M en I th L s L -e Y ta I sk LP ns
I re L vr L qs L e L hn C -g M vq LP ss
I vt L re L ev L s I cq C eg L -r FS ... 2400
2505 ... L aw F an V ks L d L s- A nn F ti LP sc
I qe C rl L rk L y L dy C th L he IR g
I ppn L et L sa I r C ts L kd L dla VP le
T ke G cc L rq L i L dd C en L qe IR 2793

L/I xx L xx L xx L x L xx C xx L xx αP **consensus**

**Figure 5.** Leucine rich repeat (LRR) COOH terminal motif of L20a. Consensus residues are present in at least half of the repeats. Numbers represent nucleotide position. "α " signifies an aliphatic amino acid and "x" signifies any amino acid.

Several of these NBS-LRR genes also contain a direct imperfect repeat sequence in the COOH region. Both L20a-5L and N have a small repeat region. Others such as L6, M, and *Rpp5*, contain an extensive region which is directly

imperfectly repeated. Still others such as Xa1 have an extensive COOH region in which a stretch of sequence is imperfectly repeated several times. Again these commonalities of basic structure with diverse specifics further support the notion that this region of the gene-product may be involved in specific recognition of some aspect of the pathogen.

We have identified a soybean gene, L20a, that directly matches the RGL NBS5 sequence that maps to a multigene family of resistance genes on linkage group F. Mapping data to date indicate that there is genetic distance between L20a and Rsv1, making this cDNA an unlikely candidate to be the Rsv1 gene. In addition *Rpv* seems to be an unlikely candidate, based on mapping data of the VP population (Saghai Maroof, unpublished). Ashfield et al. (1998) have reported that there is genetic distance between NBS5 and *Rpg1* for resistance to *Pseudomonas syringae* pv. *glycinea*, making this resistance locus an unlikely candidate. Other resistance genes in this F-linked region include Rps3 for resistance to *Phytophthora sojae* and QTLs for resistance to soybean cyst nematode (SCN (*Heterodera glycines* Ichinohe)) and root knot nematodes (RKN (*Meloidogyne javanica* and *Meloidogyne arenaria* Chitwood)) (Tamulonis et al., 1997a). Cosegregation of this clone with other disease resistance

genes in other populations, as yet has not been confirmed. It is interesting to note that major QTL's for resistance to root knot nematode have been reported on linkage group E and F (Tamulonis et al., 1997b) directly where major bands for cDNA clone L20a-5L have been mapped. These QTL's account for over 50% of the variation observed in lines segregating for resistance.

The identification of a single gene on the F-linkage group which corresponds to NBS5 suggests that RGL clones may be potentially useful in identifying functional resistance genes. Because we selected an RGL clone that was essentially single copy, we were able to locate a corresponding single copy gene. Soybean is thought to be derived from an allotetraploid progenitor (Hymowitz and Singh, 1987), therefore it is not surprising that we identified a homoeologous coding sequence on linkage group E. Shoemaker et al. (1996) have previously reported an extensive region of homoeology between the soybean E and F linkage groups. It is apparent from the sequence data that this homoeologous gene on E is not a complete resistance gene as NBS-LRR disease resistance genes are understood. Instead it seems to represent the genetic degradation of a resistance locus. That is to say, a number of mutations are evident in L33-5L which result in frame shifts and stop

codons, as compared to L20a. It is interesting to note, however, that the N-terminal region of this E-linked gene remains in tact. There is uninterrupted ORF through the entire portion of the gene that is highly conserved in other NBS-LRR resistance genes. This portion of the gene is postulated to be involved in signal transduction following triggering of the defense response. This conservation of ORF sequence could indicate some potential role for this truncated protein product.

Having identified an apparent resistance gene in soybean, the next major goal remaining is the identification of the pathogen interaction in which the NBS5 gene is involved, if any. There is the likelihood that the sequence cloned from Williams 82 represents a susceptible allele, in which case transformation experiments would be pointless. Pedigree information does not mention the presence of any resistance genes in Williams 82 that are known to be present on the F-linkage group. Further studies will have to be conducted in order to determine the nature of this apparent disease resistance gene in soybean.

**Table 1.** Regions of conservation shared by L20a-5L and other NBS/LRR type disease resistance genes. Amino acids in parentheses denote alternate residues and "x" indicates any amino acid.

| MOTIF | L20a location | Motif definition | Shared with |
|---|---|---|---|
| YD(VL)(FL)LSFR(GK)xDTR | 14-22 | | N,M,L6,RPP5 |
| IHTFx(GD)D | 42-48 | | N,M,L6,RPP5 |
| LxxAIxx(SA) | 61-68 | | N,M,L6,RPP5 |
| (VI)(VPI)(VI)(FI)Sx(NG)YAxSxWCLxxLVxI | 72-92 | | N,M,L6,RPP5 |
| (VI)(IL)P(VI)F(FY)(DM)V(ED)PSxVRxQxGx(YF) | 103-122 | | N,M,L6,RPP5 |
| WxxALxx(AI)(AG)x(LI)(SAK)G | 144-156 | | N,M,L6,RPP5 |
| (DQ)x(IFLV)(VF)G(LIR)(ED)xx(MLIV)xx(VIL)xxxL | 188-204 | | N,M,L6,RPP5 |
| (MIV)(MIVL)(GP)(ILV)xG(MIV)(GS)G(MIVLS)GK(TS)T(LIT)A(RKQ)A(VIL)(FY) | 214-233 | p-loop | N,M,L6,RPP5, I2, RPM1, Rps2, Cre3 |
| (QH)FDxx(CAI)(FW)(VLF) | 239-246 | | N,M,L6,RPP5, I2, RPM1, Rps2, Cre3 |
| QxxL(VL)(AS)E(IL)(FL)R | 262-271 | | N,M,L6,RPP5 |
| LxxK(RK)(VFLIY)(LI)(VIL)(VL)LDD(VI) | 291-303 | Kinase-2 | N,M,L6,RPP5, I2, RPM1, Rps2, Cre3 |
| GS(RK)(VIF)(MIL)(MIVLF)T(ST)R | 322-330 | Kinase-3a (putative) | N,M,L6,RPP5, I2, RPM1, Rps2, Cre3 |
| L(AS)xxx(AS)(LIFW)xL(LIF)xx(HK)AF | 349-366 | | N,M,L6,RPP5, I2, RPM1, Rps2, Cre3 |
| G(IL)PLA(ILA)xx(LIV)(GA)(SG)x(LM) | 385-397 | | N,M,L6,RPP5, I2, RPM1, Rps2, Cre3 |
| L(KR)(IVF)S(FY)DxL | 424-431 | | N,M,L6,RPP5, Rps2 |
| (KRQ)x(CIMVL)FL(DYCF)(IC)(AS)(ICL)(FY)(FPL) | 439-446 | | N,M,L6,RPP5, I2, RPM1, Rps2, Cre3 |
| L(IAV)(EDQ)(KR)(SC)(LM)(IV)x(IV) | 473-481 | | N,M,L6,RPP5 |
| Gx(VIL)xMHDL(IL)(QRE)x(ML)G(RK)(EY)I(VD) | 485-501 | | N,M,L6,RPP5 |
| PGKRSR(IL)(WT)xxE(DE)(IVG)xx(VL)Lxxx(TK)G(TS) | 509-531 | | N,M,L6,RPP5 |
| (FV)(LKV)x(ML)xxLRx(FL) | 555-564 | | N,M,L6,RPP5 |

**Literature Cited**

Aarts, M. G. M., B. L. Hekkert, E. B. Holub, J. L. Beynon, W. J. Stiekema, and A. Pereira. 1998. Identification of R-gene homologous DNA fragments genetically linked to disease resistance loci in *Arabidopsis thaliana*. Molecular Plant Microbe Interactions 11:251-258.

Ashfield, T., J. R. Danzer, D. Held, K. Clayton, P. Keim, M. A. Saghai Maroof, D. M. Webb, and R. W. Innes. 1998. *Rpg*1, a soybean gene effective against races of bacterial blight, maps to a cluster of previously identified disease resistance genes. Theoretical and Applied Genetics 96:1013-1021.

Baker, B., P. Zambryski, B. Staskawicz, and S.P. Dinesh-Kumar. 1997. Signaling in plant microbe interactions. Science 276:726-733.

Bent, A. F., B. N. Kunkel, D. Dahlbeck, K. L. Brown, R. Schmidt, J. Giraudat, J. Leung, and B. Staskawicz. 1994. Rps2 of *Arabidopsis* thaliana: a leucine-rich repeat class of plant disease resistance genes. Science 265:1856-1860.

Choi, D. W., J. Song, Y. M. Kwon, and S. G. Kim. 1996. Characterization of a cDNA encoding a proline-rich 14 kDa protein in developing cortical cells of the roots of bean (*Phaseolus vulgaris*) seedlings. Plant Molecular Biology 30:973-982.

Collins, N. C., C. A. Webb, S. Seah, J. G. Ellis, S. H. Hulbert, and A. Pryor. 1998. The isolation and mapping of disease resistance gene analogs in maize. Molecular Plant Microbe Interactions 11:968-978.

Gentzbittel, L., S. Mouzeyar, S. Badaoui, E. Mestries, F. Vear, D. T. D. Labrouhe, and P. Nicolas. 1998. Cloning of molecular markers for disease resistance in sunflower, *Helianthus annuus*, L. Theoretical and Applied Genetics 96:519-525.

Higgins, D. G. and P. M. Sharp. 1989. Fast and sensitive multiple sequence alignment on a microcomputer. CABIOS 5:151-153.

Hymowitz, T. And R. J. Singh. 1987. Taxonomy and Speciation. Agronomy 16:23-48.

Kanazin, V., L. F. Marek, and R. C. Shoemaker. 1996. Resistance gene analogs are conserved and clustered in soybean. Proceedings of the National Academy of Sciences, USA 93:11746-11750.

Lander, E. S., P. Green, J. Abrahamson, A. Barlow, J. M. Daly, S. E. Lincoln, and L. Newberg. 1987. Mapmaker: an interactive computer package for constructing primary genetic linkage maps of experimental and natural populations. Genomics 1:174-181.

Lawrence, G. J., E. J. Finnegan, M. A. Ayliffe, and J. G. Ellis. 1995. The *L*6 gene for flax rust resistance is related to the *Arabidopsis* bacterial resistance gene Rps2 and the tobacco viral resistance gene *N*. Plant Cell 7:1195-1206.

Leister, D., A. Ballvora, F. Salamini, and C. Gephardt. 1996. A PCR-based approach for isolating pathogen resistance genes from potato with potential for wide application in plants. Nature Genetics 14:421-428.

Leister, D., J. Kurth, D. A. Laurie, M. Yano, T. Sasaki, K. Devos, A. Graner, and P. Schulze-Lefert. 1998. Rapid reorganization of resistance gene homologues in cereal genomes. Proceedings of the National Academy of Sciences, USA 95:370-375.

Maughan, P. J., M. A. Saghai Maroof, and G. R. Buss. 1996. Molecular-marker analysis of seed-weight: genomic locations, gene action, and evidence of orthologous evolutionamong three legume species. Theoretical and Applied Genetics 93:574-579.

Ming-Tsan, S., T. V. Venkatesh, and R. Bodmer. 1998. Large- and small-scale preparation of bacteriophage lambda lysate and DNA. BioTechniques 25:44-46.

Ohmori, T., M. Murata, and F. Motoyoshi. 1998. Characterization of disease resistance gene-like sequences in near-isogenic lines of tomato. Theoretical and Applied Genetics 96:331-338.

Oxtoby, E., M. A. Dunn, A. Pancoro, and M. A. Hughes. 1991. Nucleotide and derived amino acid sequence of the cyanogenic beta-glucosidase (linamarase) from white clover (*Trifolium repens* L.) Plant Molecular Biology 17:209-219.

Saitou, N. and M. Nei. 1987. The neighbor joining method: A new method for reconstructing phylogenetic trees. Molecular Biological Evolution 4:406-425.

Seah, S., K. Sivasithamparam, A. Karakousis, and E. Lagudah. 1998. Cloning and characterisation of a family of disease resistance gene analogs from wheat and barley. Theoretical and Applied Genetics 97:937-945.

Shen, K. A., B. C. Meyers, M. N. I. Faridi, D. B. Chin, D. M. Stelly, and R. W. Michelmore. 1998. Resistance gene candidates identified by PCR with degenerate oligonucleotide primers map to clusters of resistance genes in lettuce. Molecular Plant Microbe Interactions 11:815-823.

Shoemaker, R. C., K. Polzin, J. Labate, J. Specht, E. C. Brummer,T. Olson, N. Young, V. Concibido, J. Wilcox, and J. P. Tamulonis. 1996. Genome duplication in soybean (*Glycine* subgenus *soja*). Genetics 144:329-338.

Speulman, E., D, Bouchez, E. B. Holub, and J. L. Beynon. 1998. Disease resistance gene homologs correlate with disease resistance loci of *Arabidopsis thaliana*. Plant Journal 14:467-474.

Tamulonis, J. P., B. M. Luzzi, R. S. Hussey, W. A. Parrott, and H. R. Boerma. 1997a. DNA markers associated with resistance to Javanese root-knot nematode in soybean. Crop Science 37:783-788.

Tamulonis, J. P., B. M. Luzzi, R. S. Hussey, W. A. Parrott, and H. R. Boerma. 1997b. DNA marker analysis of loci conferring resistance to peanut root-knot nematode in soybean. Theoretical and Applied Genetics 95:664-670.

Whitham, S., S. P. Dinesh-Kumar, D. Choi, R. Hehl, C. Corr, and B. Baker. 1994. The product of the tobacco mosaic virus resistance gene *N*: similarity to toll and the interleukin-1 receptor. Cell 78:1101-1115.

Xue, J. P., M. Lenman, A. Falk, and L. Rask. 1992. The glucosinolate-degrading enzyme myrosinase in Brassicaceae is encoded by a gene family. Plant Molecular Biology 18:387-398.

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

site. Proceedings of the National Academy of Sciences, USA 93:11751-11756.

# CHAPTER IV

# HOMOLOGY-BASED MAPPING OF THE LOCUS *Rsv1* USING A MODIFIED AFLP APPROACH

**Abstract**

In order to develop new genetic markers that might map very close to *Rsv1*, an important gene for resistance to soybean mosaic virus, we employed a marker-based system that utilizes the speed and high throughput of AFLP, but modifies it to incorporate sequence information from highly conserved regions of cloned disease resistance genes. By using labeled degenerate primers corresponding to the P-loop region of resistance genes such as *N*, *L6*, and *Rps2*, we were able to quickly amplify numerous polymorphic bands in a population segregating for resistance to *Rsv*1. Bulk segregant analysis revealed three markers that were closely linked to *Rsv*1. These markers were cloned and used as probes for RFLP analysis. The three clones mapped to within a 6 cM region surrounding *Rsv*1, the closest being 0.6 cM from Rsv1. Sequence analysis showed that all three clones contained the P-loop sequence corresponding to the degenerate primer. Further analysis indicated that one of the three clones contained open reading frame sequence which when translated was very closely related to the nucleotide binding site region of other cloned disease resistance genes.

**Introduction**

In attempting to develop high resolution maps surrounding genes of interest, it is important to generate markers that specifically target that chromosomal region of interest. A recent focus of the mapping work in our lab has been the fine mapping of the region on linkage group F surrounding *Rsv*1. *Rsv*1 is an important gene conferring resistance to soybean mosaic virus (SMV).

SMV is the most important virus infecting soybean worldwide (Thottapilly and Rossel, 1987). It has been reported wherever soybean is cultivated, and has been implicated in devastating outbreaks in several Asian countries (Irwin and Goodman, 1981). In recent years significant losses have been reported in the lower Mississippi delta, which is an important soybean growing area. Yield trials in Mississippi reported losses of up to 43% on cultivars that did not carry genetic resistance to SMV (Tolin, personal communication). A similar published study from Kentucky reported yield losses approaching 25%. (Ren et al., 1997).

The most common SMV resistance gene, *Rsv*1, has been reported from numerous genetic sources (Chen et al., 1991). The most important of these *Rsv*1 alleles derived from PI96983, confers resistance to all but the most virulent

strain, G7, of SMV. *Rsv*1 has been mapped on the soybean genome to a 5 cM-cluster of resistance genes on the F linkage group (Yu et al., 1994), which includes *Rps*3, *Rpg*1, and *Rpv*, all of which are flanked by the RFLP markers K644 and B212 (Ashfield et al., 1998; Diers et al., 1992; Saghai Maroof, unpublished). In a population of 243 $F_2$ individuals of the cross PI96983 (*Rsv*1) by Lee68 (*rsv*1), *Rsv*1 maps 0.2 cM from the RFLP marker R45 (Saghai Maroof, unpublished). Numerous RFLP, SSR, and AFLP markers have been mapped to this region.

The initial focus of fine mapping this region involved the placement of previously reported markers on a soybean map generated using this $F_2$ population segregating at the *Rsv*1 locus. Subsequent efforts have involved identifying new markers that map close to this gene. The cloning of several disease resistance genes and the subsequent identification of regions of homology has led to the development of PCR strategies that specifically detect sequences in several crops including soybean which contain these regions of conservation (Yu et al., 1996; Kanazin et al., 1996). These resistance gene-like (RGL) sequences have been mapped to numerous chromosomal regions in soybean where disease resistance genes are known to exist, including the region surrounding *Rsv1*.

Yu et al. (1996) developed a technique that consisted of designing degenerate primers from conserved sequences corresponding to two coding regions from the disease resistance genes, N (Whitham et al., 1994) and *Rps*2 (Bent et al., 1994). Specifically, these primers were designed based on the conserved P-loop and putative kinase-3a regions of the cloned genes. Using low-stringency polymerase chain reaction (PCR) they were able to amplify corresponding sequences of approximately 340 bp from soybean genomic DNA that contain similar RGL conserved sequences. Hybridization and sequence analysis showed that there are at least eleven classes of RGL sequences in soybean. The RGL clones were used as probes in a mapping population to determine their proximity to the location of known soybean disease resistance genes. Five of the eleven classes have been mapped to regions of the soybean genome where resistance genes have been identified. Two of these clones, class b (NBS5) and class j (NBS61) map to within 1 cM of *Rsv*1 on the F linkage group. Similar studies in a number of crops indicate that these RGL sequences are abundant in numerous plant genomes (Leister et al., 1996; Ohmori et al., 1998; Aarts et al., 1998; Speulman et al., 1998; Gentzbittel et al., 1998; Leister et al. 1998, Collins et al., 1998; Shen et al., 1998).

The specific identification of RGL sequences near *Rsv*1 involved PCR amplification with degenerate primers, followed by cloning of the 340 bp PCR product and characterization of individual RGL. This characterization was accomplished by hybridization to diagnostic blots containing DNA of near isogenic lines possessing resistance genes known to map to this important multigene locus on linkage group F. The success in generating these RGL clones that map close to resistance genes of interest led us to believe we could develop a strategy that might more quickly identify clones closely linked to *Rsv*1.

We propose to modify the Amplified Fragment Length Polymorphism (AFLP) procedure (Vos et al., 1995). AFLP is a PCR-based marker system where-by restriction-site and size polymorphisms are detected between lines. DNA is digested with specific restriction enzymes and then oligonucleotide adaptors are ligated to the restriction ends. These adaptors serve as the recognition site for primers during the subsequent PCR reaction. Three selective nucleotides at the 3' end of the primer ensure amplification of only a small subset of the digested and ligated genomic fragments. These fragments are labeled and separated on a polyacrylamide gel in order to detect polymorphic bands.

The objectives of this study are to develop and test a PCR-based strategy that combines the speed, efficiency, and high throughput of AFLP with the resistance-gene-directed approach of NBS to generate markers tightly linked to *Rsv*1.

**Materials and Methods**

Two mapping populations were used in the present study. The $F_2$ population PI96983 (*Rsv*1) X Lee68 (*rsv*1) (PL population) was used for bulk segregant analysis and fine mapping of the *Rsv*1 locus. This population contains 243 $F_{2:3}$ individuals and has been saturated with molecular markers surrounding the locus, *Rsv*1. The $F_2$ population V71-370 X PI407162 (VP population) which contains over 300 mapped loci covering all 20 linkage groups (Saghai Maroof, unpublished) was used for determining the location of markers not linked to *Rsv*1.

Young trifoliate leaf tissue from greenhouse-grown plants of 12 to 15 $F_{2:3}$ individuals from both populations were collected. DNA was isolated by freeze-drying followed by CTAB extraction as described (Saghai Maroof et al., 1984). DNA from twelve homozygous resistant and susceptible lines of the PL population were pooled to form bulk resistant and bulk susceptible samples. These DNA samples along with parental DNA samples PI96983 and Lee68 were

digested with restriction enzymes *Eco*RI and *Mse*I, and adaptor sequences were ligated to the restriction ends as described by Vos et al. (1995). An initial round of PCR was conducted using *Mse*I and *Eco*RI +1 selective primers, to amplify a subset of the digested and ligated fragments. Thirty cycles of a 25 uL PCR reaction using 1X buffer, 0.9 mM MgCl2, 0.6 uM of each primer, 0.25 mM dNTP, 0.5 U of *Taq* polymerase and 150 ng of template was conducted with denaturation at 94°C for 60 sec., primer annealing at 60°C for 30 sec., and primer extension at 72°C for 30 sec. An aliquot of this reaction equivalent to 25ng was then used in a second round of amplification. The *Eco*RI +1 primer used in the first reaction and a $^{32}$P-end labeled degenerate primer (corresponding to the P-loop region of resistance genes, *N*, *Rps*2, and *L*6) were used in the second round PCR reaction. The degenerate primer sequence (5' GGA ATG GGN GGN GTN GGN AAR AC 3') is the same as that used by Yu et al. (1996). PCR components were essentially the same as the previous reaction with the exception that 0.15 uM of labeled and 0.6 uM of unlabeled primer were used in combination with 0.75 uM of adapter specific +1 primer. This second round of PCR was conducted using touchdown conditions beginning at 65°C annealing and reducing by 1°C per cycle to 56°C,

followed by 26 cycles at 56°C. All other cycling conditions were as the previous reaction.

Second round PCR products were separated on a 7 M urea 6.5% polyacrylamide gel for 2.5 h at 60 W. The gel was then transferred to 3MM paper, covered with plastic wrap, and exposed to Kodak (New Haven, CT) Xomat film for 12 to 18 hours.

PCR products which appeared to be linked based on bulk segregant analysis were excised from the gel by elution in 100 µL of water incubated in a boiling bath as described by Upender et al. (1995). Eluate was PCR amplified using the same primers that generated the polymorphic product under the same or less stringent conditions. The PCR product was then cloned into the pCNTR shuttle vector using the General Contractor Cloning Kit (5 prime, 3 prime, Boulder, CO). Cloned inserts were amplified with the original +1 primer and labeled degenerate primer again under the same second round PCR conditions. These labeled fragments were then run on a 7M urea, 6.5% polyacrylamide gel, along-side labeled parental samples generated under the same conditions to determine those clones that contained the proper size insert based on BSA. Tentatively confirmed positive inserts were amplified for use as RFLP markers.

The Southern hybridization procedure was as described (Yu et al., 1994). Briefly, 8 µg of parental DNA was digested with enzymes *Dra*I, *Eco*RI, *Eco*RV, *Hin*dIII, *Xba*I, and *Taq*I, according to manufacturers protocols (Gibco-BRL, Rockville, MD). Digested DNA was then separated on a 1% agarose gel at 70 to 90 mAmps for 14 to 16 hours. The DNA was transferred to Hybond nylon membrane (Amersham, Piscataway, NJ) by Southern blotting with 0.4 N NaOH buffer. Screening blots were hybridized with $\alpha^{32}$PdCTP, random primer labeled probe. Hybridizing bands were visualized by autoradiography on Kodak (New Haven, CT) Xomat film. Polymorphic bands were mapped in the VP and PL population. Multiple linkage analysis was determined using MAP-MAKER software at LOD 3.0 (Lander et al., 1987).

In addition, inserts were sequenced using an ABI377 DNA sequencer. Plasmid template was prepared using standard alkaline-lysis followed by purification using QiaexII (Qiagen, Valencia, CA). Dye-terminator cycle sequencing was done based on manufacturer’s protocols. Sequence assembly and analysis was conducted using Lasergene software from DNAstar (Madison, WI).

**Results**

An initial test of the homology-based AFLP method on several parental lines indicated that *Eco*+1 primers combined with the degenerate P-loop primer resulted in a better spread of bands on the gel than did the *Mse*+1/P-loop combination. Twelve *Eco*+1/P-loop primer combinations were tested with bulk and parental DNA samples of the PL population. A total of 23 polymorphisms were observed among the total 720 bands visualized. Three dominant markers, R12 (171 bp), R13 (261 bp), and R14 (330 bp), were identified which appeared to be linked to *Rsv*1 (Fig. 1). Amplified product from the parental and bulk samples were excised from the gel and cloned for further analysis. In addition to these putatively linked bands, several other bands were selected for excision and cloning. These clones were designated A, C, X, and Z.

Plasmid from several colonies of each positive clone was extracted and an AFLP reaction was run using the same primer combination that generated the original polymorphic fragment. These samples were then run along side parental samples on an acrylamide gel to identify those clones which were of the proper size. Insert from these putatively positive clones was then hybridized to diagnostic blots to identify mappable polymorphisms. Banding patterns of proper

sized inserts indicated that each group of positive clones from a putatively linked AFLP contained colonies of two distinct soybean sequences. That is to say, R12 clones of

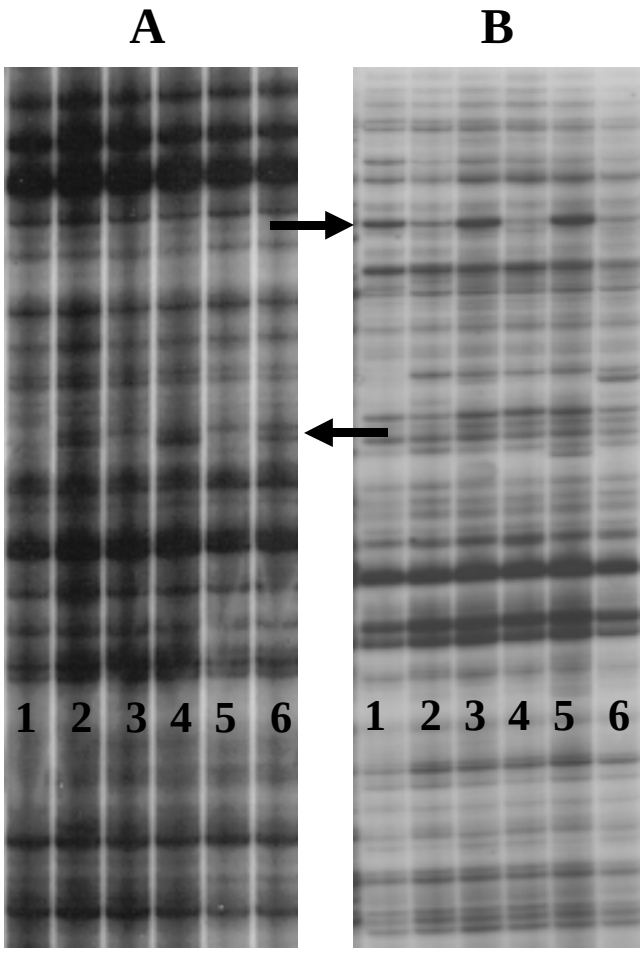



**Figure 1.** Homology-based AFLP of PI96983, Lee68 and bulk resistant and susceptible $F_2$s generated from the *Eco*+C/*Mse*+T pool with selective primers *Eco*+C and the labeled degenerate P-loop primer (A) and *Eco*+T/*Mse*+T pool with selective primers *Eco*+T and the labeled degenerate P-loop primer (B). Lane 1 is Lee68, lane 2 is PI96983, lanes 3 and 5 are susceptible bulks, and lanes 4 and 6 are resistant bulks. The arrows show the *Rsv*1 linked polymorphisms, R12 (A) and R14 (B). For R12 the band is present in PI96983 and resistant bulks. For R14 the band is present in Lee68 and susceptible bulks.

proper insert size showed either one or the other of two distinct banding patterns. The same was true of R13 and R14.

One representative of each clone-type was chosen, designated R12.4, R12.10, R13.48, R13.61, R14.111, and R14.113. These clones were mapped in the VP population containing over 300 mapped loci. In each case one type of clone represented a soybean sequence closely linked to Rsv1 and the other clone type represented an unrelated sequence not closely linked to Rsv1. These unlinked sequences likely represented “contaminant” amplified sequences that migrated to a similar location on the acrylamide gel. In addition, the unlinked clones which were excised were mapped to determine their genetic location.

The three putatively linked clones, R12.4, R13.48, and R14.113, were mapped with respect to *Rsv*1 in the PL population which contains 243 $F_2$ individuals segregating for *Rsv*1. Two of the markers, R13.48 and R14.113, map close to previously identified NBS markers at distances of 0.6 and 1.6 cM from *Rsv*1. R14.113 cosegregates with the previously identified NBS5 clone in the PL population. The third marker, R12.4, maps to the other side of *Rsv*1, 3.9 cM from the gene (Fig. 2).

In addition the three markers that are not linked to *Rsv*1, R12.10, R13.72, and R14.111, as well as the clones A10, C3, X20, and Y37 were mapped. All seven of these markers were mapped in the VP population. Several of these markers map to more than one locus. Map locations of these markers are as listed in Table 1.

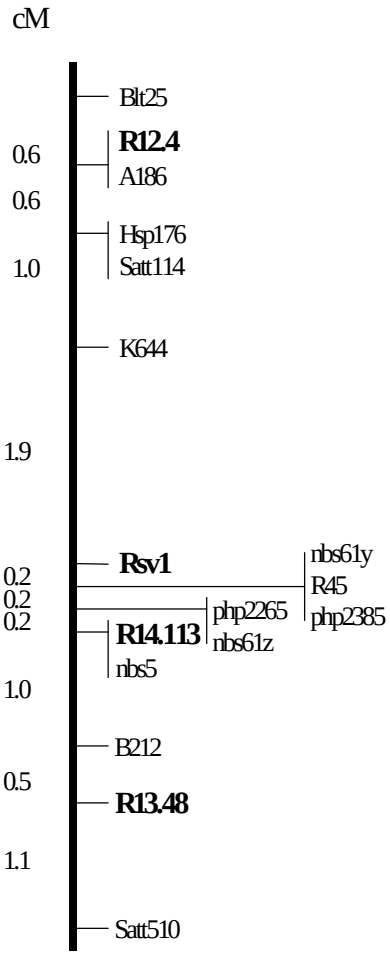


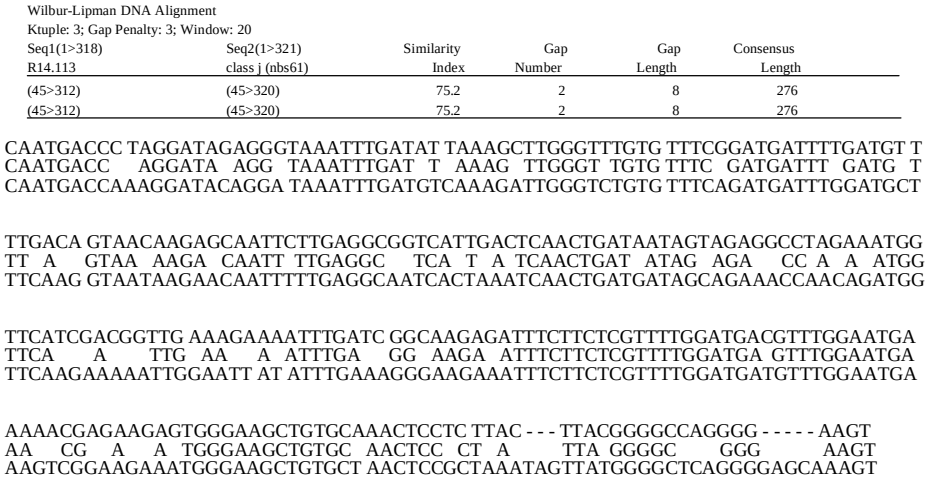

**Figure 2.** Map of linkage group F surrounding the SMV resistance locus, *Rsv*1, showing the location of three homology-based AFLPs (in bold) at this genomic location. Map data are calculated based on a 243 individual F2 population of the cross PI96983 X Lee68.

**Table 1.** Additional homology-based markers converted and mapped as RFLPs in the population V71-370 X PI407162 showing the marker name and the soybean linkage group where each maps. Names followed by a letter indicate multiple loci mapped with a single probe.

| MARKER | LINKAGE GROUP |
|---|---|
| R12.10a | A1 |
| R12.10b | L |
| R12.10c | F |
| R13.61a | H |
| R13.61b | D2 |
| R13.61c | D1a |
| R14.111 | J |
| A10 | H |
| C3a | C1 |
| C3b | C2 |
| X20 | G |
| Y37a | L |
| Y37b | C1 |
| Y37c | K |

Sequence analysis of all ten of the homology-based AFLP clones identified, showed that each contains the highly conserved P-loop region that corresponds to the labeled

degenerate primer used to amplify the fragments. Alignment of these sequences with previously identified NBS sequences showed that only one of the ten clones has motifs consistent with disease resistance related sequences. This clone, R14.113, is 312 nucleotides in length excluding the adaptor sequence which had been ligated to the fragment. This clone maps to within 0.6 cM of Rsv1, and cosegregates with another NBS clone, NBS5. Alignment of R14.113 with other soybean



Wilbur-Lipman DNA Alignment
Ktuple: 3; Gap Penalty: 3; Window: 20

| Seq1(1>318) R14.113 | Seq2(1>321) class j (nbs61) | Similarity Index | Gap Number | Gap Length | Consensus Length |
|---|---|---|---|---|---|
| (45>312) | (45>320) | 75.2 | 2 | 8 | 276 |
| (45>312) | (45>320) | 75.2 | 2 | 8 | 276 |

CAATGACCC TAGGATAGAGGGTAAATTTGATAT TAAAGCTTGGGTTTGTG TTTCGGATGATTTTGATGT T
CAATGACC AGGATA AGG TAAATTTGAT T AAAG TTGGGT TGTG TTTC GATGATTT GATG T
CAATGACCAAAGGATACAGGA TAAATTTGATGTCAAAGATTGGGTCTGTG TTTCAGATGATTTGGATGCT

TTGACA GTAACAAGAGCAATTCTTGAGGCGGTCATTGACTCAACTGATAATAGTAGAGGCCTAGAAATGG
TT A GTAA AAGA CAATT TTGAGGC TCA T A TCAACTGAT ATAG AGA CC A A ATGG
TTCAAG GTAATAAGAACAATTTTTGAGGCAATCACTAAATCAACTGATGATAGCAGAAACCAACAGATGG

TTCATCGACGGTTG AAAGAAAATTTGATC GGCAAGAGATTTCTTCTCGTTTTGGATGACGTTTGGAATGA
TTCA A TTG AA A ATTTGA GG AAGA ATTTCTTCTCGTTTTGGATGA GTTTGGAATGA
TTCAAGAAAAATTGGAATT AT ATTTGAAAGGGAAGAAATTTCTTCTCGTTTTGGATGATGTTTGGAATGA

AAAACGAGAAGAGTGGGAAGCTGTGCAAACTCCTC TTAC - - - TTACGGGGCCAGGGG - - - - - AAGT
AA CG A A TGGGAAGCTGTGC AACTCC CT A TTA GGGGC GGG AAGT
AAGTCGGAAGAAATGGGAAGCTGTGCT AACTCCGCTAAATAGTTATGGGGCTCAGGGGAGCAAAGT

**Figure 3.** Wilbur-Lipman pairwise nucleotide alignment of the homology-based AFLP clone, R14.113 (top row), with the class j clone, NBS 61(bottom row). The middle row indicates nucleotides that are identical between the two clones. Dashes represent gaps introduced to better reflect the alignment of the two clones.

NBS sequences showed that it most closely resembles NBS61, sharing 75% similarity based on Wilbur-Lipman pairwise alignment (Wilbur and Lipman, 1983) (Fig. 3).

One of the mapped clones, A10, was shown to be a simple sequence repeat marker by sequence analysis. This clone maps to the soybean linkage group H 12.4 cM from the marker pK14.

**Discussion**

When fine mapping a gene of agronomic importance such as disease resistance loci, it is a useful strategy to enhance the detection of markers which specifically target the gene of interest. In addition to bulk segregant analysis and traditional AFLP, we have employed a novel homology-based AFLP approach which has quickly identified three tightly linked markers to *Rsv*1, an important soybean resistance gene.

The presence of numerous P-loop type sequences facilitated the generation of these molecular markers. In this study we used a modified AFLP approach where-by only those sequences that contained a P-loop like region near an *Eco*RI restriction site were amplified and visualized. In the initial digestion and ligation steps, approximately 87.5% of the soybean genome is eliminated because *Mse*/*Mse*

fragments are not detected in the proceeding steps. In this study we evaluated 12 of the 16 possible primer combinations using only the *Eco*+1 primer. On average we obtained 60 bands per reaction which represents no less than 60 P-loop sequences per gel. This does not include amplification products that migrate to the same point on the gel. Based on these observations we estimate there are no less than 15000 of these type of P-loop sequences in the soybean genome. This is based on the percentage of the genome sampled, the mean number of products observed per sampling, the total number of sampling combinations, and the fact that only one of the two DNA strands is detected by labeled primer annealing per fragment.

Of the 10 clones evaluated in this study, all contained the conserved p-loop sequence which is present in a large class of disease resistance genes. Of the three markers linked to *Rsv1*, one, R14.113, contains several motifs identical to other cloned NBS sequences. This clone cosegregates with another RGL, NBS5. Interestingly, overall sequence similarity to NBS5 is relatively low at 27.7%. Conversely, R14.113 is very similar structurally to a second RGL clone, NBS61, which represents a multigene family mapping to the *Rsv*1 region. These two clones have a similarity of 75.2%. Two NBS61 loci are located 0.2 and 0.4

cM from this clone suggesting that all of these loci may be related sequences of a multi-gene family near *Rsv*1. Like NBS61, R14.113 shows multiple bands on an RFLP blot. There are, however, no apparent commonalities between the banding patterns of these two clones.

The PCR-based marker system developed in this study, specifically amplifies sequences that contain motifs known to occur in a number of hypersensitive-response type resistance genes. These NBS-LRR genes all contain P-loop, kinase-2, and putative kinase-3a motifs. Using a single restriction enzyme combination and only a few primer combinations we were able to quickly identify three markers containing P-loop sequences that tightly flank our resistance gene of interest, *Rsv*1. Identification of additional markers closely linked to other resistance genes, particularly in those regions known to contain RGL clones, or other NBS-LRR type resistance genes is certainly plausible using this technique. In addition, it is certainly possible to modify this technique by using other degenerate primers which correspond to other conserved motifs from NBS-LRR and other classes of resistance genes.

While this technique may have wide applicability with regard to mapping important crop resistance genes, it is not applicable in all situations. We are currently mapping a

second SMV resistance gene in soybean. This gene, Rsv4, has a different mode of action than the hypersensitive response type of resistance which is characteristic of all the NBS-LRR type cloned resistance genes in addition to *Rsv*1 (Ma et al., 1995). Attempts to identify any RGL or P-loop related sequences, including homology-based AFLP, that map close to this resistance gene have been unsuccessful. *Rsv*4 maps to a region of the chromosome where no hypersensitive response resistance genes have been identified. Thus, the likelihood of identifying these homology-based sequences in this region is somewhat lower than other regions where RGL sequences and hypersensitive response genes have been identified.

**Literature Cited**

Aarts, M. G. M., B. L. Hekkert, E. B. Holub, J. L. Beynon, W. J. Stiekema, and A. Pereira. 1998. Identification of R-gene homologous DNA fragments genetically linked to disease resistance loci in *Arabidopsis thaliana*. Molecular Plant Microbe Interactions 11:251-258.

Ashfield, T., J. R. Danzer, D. Held, K. Clayton, P. Keim, M. A. Saghai Maroof, D. M. Webb, and R. W. Innes. 1998. Rpg1, a soybean gene effective against races of bacterial blight, maps to a cluster of previously identified disease resistance genes. Theoretical and Applied Genetics 96:1013-1021.

Bent, A. F., B. N. Kunkel, D. Dahlbeck, K. L. Brown, R. Schmidt, J. Giraudat, J. Leung, and B. Staskawicz. 1994. Rps2 of *Arabidopsis* thaliana: a leucine-rich repeat class of plant disease resistance genes. Science 265:1856-1860.

Chen, P., G. R. Buss, C. W. Roane, and S. A. Tolin. 1991. Allelism among genes for resistance to soybean mosaic virus in strain-differential soybean cultivars. Crop Science 31:305-309.

Collins, N. C., C. A. Webb, S. Seah, J. G. Ellis, S. H. Hulbert, and A. Pryor. 1998. The isolation and mapping of disease resistance gene analogs in maize. Molecular Plant Microbe Interactions 11:968-978.

Diers, B. W., L. Mansur, J. Imsande, and R. C. Shoemaker. 1992. Mapping phytophthora resistance loci in soybean with restriction fragment length polymorphism markers. Crop Science 32:377-383.

Gentzbittel, L., S. Mouzeyar, S. Badaoui, E. Mestries, F. Vear, D. T. D. Labrouhe, and P. Nicolas. 1998. Cloning of molecular markers for disease resistance in sunflower, *Helianthus annuus*, L. Theoretical and Applied Genetics 96:519-525.

Irwin, M. E. and R. M. Goodman. 1981. Ecology and control of soybean mosaic virus. pp. 181-220 *in* Maramorosch, K. and K. F. Harris (eds.) Plant diseases and vectors: ecology and epidemiology. Academic Press, New York.

Kanazin, V., L. F. Marek, and R. C. Shoemaker. 1996. Resistance gene analogs are conserved and clustered in

soybean. Proceedings of the National Academy of Sciences, USA 93:11746-11750.

Lander, E. S., P. Green, J. Abrahamson, A. Barlow, J. M. Daly, S. E. Lincoln, and L. Newberg. 1987. Mapmaker: an interactive computer package for constructing primary genetic linkage maps of experimental and natural populations. Genomics 1:174-181.

Leister, D., J. Kurth, D. A. Laurie, M. Yano, T. Sasaki, K. Devos, A. Graner, and P. Schulze-Lefert. 1998. Rapid reorganization of resistance gene homologues in cereal genomes. Proceedings of the National Academy of Sciences, USA 95:370-375.

Leister, D., A. Ballvora, F. Salamini, and C. Gephardt. 1996. A PCR-based approach for isolating pathogen resistance genes from potato with potential for wide application in plants. Nature Genetics 14:421-428.

Ma, G., P. Chen, G. R. Buss, and S. A. Tolin. 1995. Genetic characteristics of two genes for resistance to soybean mosaic virus in PI486355 soybean. Theoretical and Applied Genetics 91:907-914.

Ohmori, T., M. Murata, and F. Motoyoshi. 1998. Characterization of disease resistance gene-like sequences in near-isogenic lines of tomato. Theoretical and Applied Genetics 96:331-338.

Ren, Q. T., W. Pfeiffer, and S. A. Ghabrial. 1997. Soybean mosaic virus incidence level and infection time: Interaction effects on soybean. Crop Science 37:1706-1711.

Saghai Maroof, M. A., K. M. Soliman, R. A. Jorgensen, and R. W. Allard. 1984. Ribosomal DNA spacer length polymorphism in barley: Mendelian inheritance, chromosomal location, and population dynamics. Proceedings of the National Academy of Sciences, USA 81:8014-8018.

Shen, K. A., B. C. Meyers, M. N. I. Faridi, D. B. Chin, D. M. Stelly, and R. W. Michelmore. 1998. Resistance gene candidates indentified by PCR with degenerate oligonucleotide primers map to clusters of resistnce genes in lettuce. Molecular Plant Microbe Interactions 11:815-823.

Speulman, E., D, Bouchez, E. B. Holub, and J. L. Beynon. 1998. Disease resistance gene homologs correlate with disease resistance loci of *Arabidopsis* thaliana. Plant Journal 14:467-474.

Thottapilly, G. and H. W. Rossel. 1987. Viruses affecting soybean in Singh, S. R., K. O. Rachie, and K. E. Dashiell (eds.). Soybeans for the tropics: research, production, utilization. John Wiley and Sons Ltd., Chichester, U.K. pp.53-68.

Upender, M., L. Raj, and M. Weir. 1995. Rapid method for the elution and analysis of PCR products separated on high resolution acrylamide gels. Biotechniques 18:33-34.

Vos, P., R. Hogers, M. Bleeker, M. Reijans, T. van de Lee, M. Hornes, A. Frijters, J. Pot, J. Peleman, M. Kuiper, and M. Zabeau. 1995. AFLP: a new technique for DNA fingerprinting. Nucleic Acids Research 23:4407-4414.

Whitham, S., S. P. Dinesh-Kumar, D. Choi, R. Hehl, C. Corr, and B. Baker. 1994. The product of the tobacco mosaic virus resistance gene *N*: similarity to toll and the interleukin-1 receptor. Cell 78:1101-1115.

Wilbur, W. J., and D. J. Lipman. 1983. Rapid similarity searches of nucleic acid and protein databases. Proceedings of the National Academy of Sciences, USA 80:726-730.

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

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

## Vita

Alec Hayes was born on January 4, 1969, in Richmond, VA, to Harlan and Betty Hayes. His entire secondary education was undertaken in Amelia County, VA. He graduated from Amelia County High School in the spring of 1987. For two years he studied at Longwood College in Farmville, VA. In the fall of 1989 he transferred to the College of William and Mary in Williamsburg, VA, where he completed his undergraduate degree, obtaining a B.S. in biology with an emphasis in genetics.

After graduating in the spring of 1991, he took a job at Chippenham Medical Center in Richmond, VA. There he served in the pathology and cytotechnology department, learning the techniques of preparing and staining tissue samples for cytogenetic evaluation.

In the summer of 1993, he returned to his education, by pursuing his Master's degree under the guidance of Dr. Carol Wilkinson at Virginia Tech and The Southern Piedmont Agricultural Research and Extension Center. There he studied the principles of classical genetics as it relates to plant breeding. In addition to learning the basics of crossing and selection, he conducted a study to evaluate tobacco germplasm for natural resistance to tobacco cyst nematode, an important pathogen of the southern Piedmont. He received his Master's degree in the spring of 1995.

Directly following this he began working in the laboratory of Dr. M. A. Saghai Maroof, in the department of Crop and Soil Environmental Sciences, at Virginia Tech. There he learned and practiced the skills of genetic analysis utilizing various molecular techniques. The emphasis of his study in this capacity has been genomic analysis of disease resistance in soybean. In addition to his laboratory work, he has served for three years as a teacher for the laboratory section of the introductory soils class at Virginia Tech.

Alec was inducted into the Gamma Sigma Delta agricultural honor society and the Sigma Phi biological honor society.