To the Graduate Council:

I am submitting herewith a thesis written by Shane Evan Abbitt entitled "Analysis of soybean non-nodulation mutants by DNA Amplification Fingerprinting and Bulked Segregant Analysis". I have examined the final copy of this thesis for form and content and recommend that it be accepted in partial fulfillment of the requirements for the degree of Master of Science, with a major in Life Sciences.

Dr. Peter Gresshoff, Major Professor

We have read this thesis and

recommend its acceptance:

Accepted for the Council:

Associate Vice Chancellor and

Dean of the Graduate School

Analysis of soybean non-nodulation mutants by DNA amplification fingerprinting and bulked segregant analysis

A thesis

presented for the

Master of Science

Degree

The University of Tennessee, Knoxville

Shane Abbitt

August, 1996

Acknowledgments

There are many people who have been instrumental in my growth as a scientist , as well as an individual, during my time at the University of Tennessee. I consider it a luxury to have spent time in the Plant Molecular Genetics lab with such a diversity of research as well as people. It has been a great benefit to listen to these great minds at work. I would especially like to thank Janice Crockett, and Debbie Landau-Ellis for all of their help and patience. Without these two ladies, life in the PMG lab would be much more difficult. It has also been a tremendous learning experience to work under Dr. Peter Gresshoff, who is the catalyst of everything in the PMG lab. I truly believe that it is impossible to come away from a conversation with him and not learn something new. I would also like to thank my committee members, Gustavo Caetano-Anollés, Gary Stacey, and Don Dougall for their support and guidance. My family has also been a great help. I would like to thank my wife Tara, for patience and encouragement, my mother-in-law and father-in-law for a home away from home, and most of all, my parents for many years of support and encouragement.

Abstract

In this study, two techniques were combined in order to create a more efficient means to isolate molecular markers for two unlinked non-nodulation loci in soybean. These two techniques had not previously been used in unison but have been shown to have characteristics sufficient to meet the goal.

These techniques are: 1) DNA Amplification Fingerprinting (DAF) and 2) Bulked Segregant Analysis (BSA). DAF is a technique that employs the use of arbitrary primers to amplify many regions of genomic DNA. BSA is a technique of pooling individuals from a population into bulks based on a phenotypic trait that can be scored. In using these techniques together, BSA was supposed to isolate molecular markers in one bulk so that DAF could be used to detect that difference.

The results were not as expected and no linked markers were isolated. The focus then turned to analyzing the experimental design to find exactly why the anticipated results were not obtained. It was found that the nature of the genome contributed most of the problems. The fact that these loci are in regions of conservation and that the soybean genome is most likely duplicated are major contributions to the pitfalls of the project design. Competition during amplification of the DNA also seems to have some effects.

Suggestions to make the next effort in isolating markers more successful included the removal of the BSA technique from the project design.

Table of contents

Chapter		page
1. Introduction	.....	1
I. Basics of nodulation	.....	1
II. Non-nodulation mutants	.....	1
a. nod49	.....	7
b. nod139	.....	7
III. Genetic markers and their detection	.....	8
IV. Bulked segregant analysis (BSA)	.....	10
V. Project goal	.....	12
2. Materials and methods	.....	16
I. Plant tissue collection	.....	16
a. nod49	.....	16
b. nod139	.....	16
II. DNA extraction and isolation	.....	19
III. DNA quantification	.....	20
IV. Bulking of DNA samples	.....	20
V. DNA amplification fingerprinting (DAF)	.....	21
VI. DNA separation and silver staining	.....	23
VII. Scoring of DAF products	.....	24
3. Results	.....	26
I. Progeny screen of nod139	.....	26
II. Parental screening	.....	26
III. Mini-hairpin screening	.....	27
IV. Bulk screening	.....	29

a. nod49 29

b. nod139 36

V. Competition experiments 41

4. Discussion 51

I. Primer screening 51

II. Competition experiments 52

III. Band competition theories 54

IV. The lack of closely linked markers..... 58

V. Future directions 61

VI. Lessons learned 64

Literature cited 66

Vita 75

List of Tables

Table	Title	page
(2-1)	G43 generation genotypes	17
(2-2)	M70 generation genotypes	18
(2-3)	Primer sequences	22
(3-1)	Mini-hairpin parental screen of *G. soja* and nod49	27
(3-2)	Summary of mini-hairpin parental screen data	28
(3-3)	Primer screening of nod49 and nod139 bulks	30
(3-4)	nod49 polymorphic bands data	33
(3-5)	Polymorphic band distribution from nod49 bulks and parents	34
(3-6)	nod49 individual screen results	35
(3-7)	nod139 polymorphic band data	38
(3-8)	Polymorphic band data from nod139 bulks and parents	39
(3-9)	nod139 individual screen results	40
(3-10)	Data from competition experiment I (Bermudagrass vs. *G. soja*)	42
(3-11)	Data from competition experiment II (*G. soja* vs. nod49)	47

List of Figures

Figure	Title	page
(1-1)	Actions of nodulation mutants during the stages of nodulation	5
(1-2)	Roots of cv. 'Bragg, nod139, and nod49	6
(1-3)	Strategy involved in Bulked Segregant Analysis	13
(3-1)	nod49 polymorphic band profiles	32
(3-2)	nod139 polymorphic band profiles	37
(3-3)	Bermudagrass vs. *G. soja* template competition	43
(3-4)	Cumulative percentage of total bands lost in Bermudagrass vs. *G. soja* template competition	44
(3-5)	Band competition experiment I (Bermudagrass vs. *G. soja*)	45
(3-6)	*G. soja* vs. nod49 template competition	48
(3-7)	Cumulative percentage of total bands lost in *G. soja* vs. nod49 template competition	49
(3-8)	Band competition experiment II (*G. soja* vs. nod49)	50
(4-1)	Agarose gel electrophoresis of templates	57
(4-2)	Demonstration of amplification potential effects	59
(4-3)	The effects of genome duplication	62

List of abbreviations

AFLP	Amplified Fragment Polymorphism
AP-PCR	Arbitrarily Primed-Polymerase Chain Reaction
BSA	Bulked Segregant Analysis
BP	Base Pairs
DAF	DNA Amplification Fingerprinting
HP	Mini-Hairpin
PCR	Polymerase Chain Reaction
RAPD	Random Amplified Polymorphic DNA
RFLP	Restriction Fragment Length Polymorphism
SDI	Shoot Derived Inhibitor
STR	Short Tandem Repeat
YAC	Yeast Artificial Chromosome

Chapter 1
Introduction

I. Basics of legume nodulation

Nitrogen is one of the major elements controlling the growth of plants. It is largely used in amino acids, as well as the purine and pyrimidine bases that make up DNA and RNA. When the availability of nitrogen in the soil is low, plants have to turn to alternative sources. Legumes have attained the ability to gain fixed nitrogen through a symbiotic interaction with certain strains of bacteria known as *Rhizobium* or *Bradyrhizobium*. The *Rhizobium* induced infection and symbiosis results in an exchange of fixed carbon for fixed nitrogen. It is a complex process that results in the formation of nodules on the root and fixation of nitrogen by the *Rhizobium* inside these nodules.

While studying the effects of nitrogen nutrition on the yield on different plants, it was discovered that legumes reacted differently than other plant species (Hellriegel and Wilfarth, 1888). When nitrates were not supplied, most plants had a yield approaching zero. Legumes, however, had a surprisingly high yield. The inevitable conclusion was that legumes had found an alternate nitrogen source which had to be atmospheric nitrogen gas. In another series of experiments, sterilized legume seeds were grown under sterile conditions without nitrate (Hellriegel and Wilfarth, 1889). They grew poorly, like other plant species, until inoculated with field soils when they assumed their growth advantage. These experiments demonstrated that legumes can access atmospheric nitrogen through an interaction with soil bacteria (Quispel, 1988)

This symbiotic exchange of nutrients takes place in a bacterial induced root structure called a nodule. The formation of nodules is an intricate system. This process occurs in three

steps: 1) Nodule initiation, 2) Infection by *Rhizobium* and, 3) Nitrogen fixation (Gresshoff and Delves, 1986). Initiation begins with the production of exudates from the roots. These exudates are flavonoid compounds which attract *Rhizobium* to the root hairs as well as induce bacterial gene expression (Peters et al, 1986, Redmond et al, 1986, Kosslak et al, 1987). Nod factors are lipo-oligosaccharides produced by the bacteria that induce root hair deformation (Lerouge et al, 1990) and root hair curling (Relic et al, 1993). They induce pre-infection structures (van Brussel et al, 1992) and nodules (Truchet et al, 1991). Nod factors also permit *Rhizobium* to penetrate legume roots through the infection threads (Relic et al, 1994). During the infection stage, infection threads grow from the entrapped bacteria to the root cortex, giving the bacteria access to the plant cytoplasm. The bacteria then begin to fix nitrogen and the exchange of nutrients occurs. The legume-*Rhizobium* interaction is very specific, the host infected only by certain bacteria and the bacteria infecting only certain hosts. Some exceptions, such as *Rhizobium* strain NGR234, can nodulate a variety of legumes (Caetano-Anollés and Gresshoff, 1991c). The specificity is controlled by bacterial genes called host-specificity of nodulation (*hsn*) genes located on the symbiotic plasmid (Kondorosi et al, 1984). *hsn* genes modify the "common" Nod factor into a "host-specific" by the addition of various substiuents. For example, the *hsn* gene, *nodH*, is responsible for sulfation of the Nod factor (Carlson et al, 1994).

The symbiotic interaction is controlled by genes located in both the host and the microsymbiont. The *nif* genes are bacterial genes responsible for the pathways that lead to nitrogen fixation. These genes code for proteins such as nitrogenase. Proteins produced exclusively by the host, called Nodulins, are responsible for actions such as oxygen transport and cell wall architecture (Caetano-Anollés and Gresshoff, 1991c). Evidence for host symbiosis genes came from the discovery of non-nitrogen-fixing nodules, formed in the absence of *Rhizobium*, in alfalfa (Truchet et al, 1989). The presence of these nodules

point to the legume as the location of the genes responsible for nodule structure (Caetano-Anollés and Gresshoff, 1991c). Many other host genes are involved in nodule formation through regulatory control pathways (Gresshoff, 1993).

Host genes are responsible for the control mechanisms of nodule formation. Pathways responsible for this control are: 1) nitrate inhibition and, 2) autoregulation. Nitrate exhibits an inhibitory effect on nodulation in a localized and nonsytemic manner (Carroll and Gresshoff, 1983). In split root experiments where one half of the roots were exposed to *Bradyrhizobium* and nitrate and the other half exposed to only *Bradyrhizobium*, nodule development of the nitrate-exposed roots was inhibited. Normal nodulation continued on the nitrate-free roots, however, nitrogenase activity was inhibited on both sides (Gresshoff, 1993). Autoregulation is an internal, homeostatic control of the amount of nodule tissue produced. It acts by inhibiting infection-susceptible roots from further nodule development after the establishment of prior infections. This controls the amount of nodule meristem present in the legume, and therefore, the amount of "resources' spent on nodule formation (Caetano-Anollés and Gresshoff, 1991c). In soybean experiments, nodules were removed from the root. New nodules then developed in the original region of nodulation (Caetano-Anollés et al, 1991e, Caetano-Anollés and Gresshoff, 1991a). These data suggest that autoregulation has a very early effect on nodule formation and that the effect is maintained by the original nodules (Gresshoff, 1993).

There are many plant genes involved in the control of nodulation and nitrogen fixation, most have yet to be discovered. Steps to isolating these genes include the use of nodulation mutants. Mutants nts382 and nts1007, produced by ethyl methane sulphonate (EMS) mutagenesis of *G. max* (L.) Merr cv 'Bragg', exhibit a supernodulating phenotype. They have been shown to have up to 30 times as many nodules as 'Bragg'. The "nts"

designation refers to "nitrate tolerant symbiosis", meaning these mutants form an extreme amount of nodules, in the presence of nitrate. Both nitrate inhibition and autoregulation are suppressed in these mutants (Carroll et al, 1985). Approach graft experiments were performed by grafting one vertically sectioned nts382 plant to a vertically sectioned 'Bragg' plant to create a "whole" plant. Results of these experiments showed that the supernodulating phenotype results from the mutation of an autoregulatory signal produced in the shoot of the plant. This signal, termed a shoot derived inhibitor (SDI), is responsible for systemic suppression of nodulation (figure 1-1) (Caetano-Anollés and Gresshoff, 1990).

Non-nodulation mutants are also important in the isolation of the genes (figure 1-2). Three soybean non-nodulation mutants have been isolated, through EMS and M_2 family selection, that define two different loci. They include nod49, nod772, and nod139, (Carroll et al, 1986). Nod49 and nod772 were shown to be allelic, but nod139 appears to be unlinked (Mathews et al, in review). Naturally occurring soybean nodulation mutants Rj_2 (Caldwell, 1966), Rj_3 (Vest, 1970), and Rj_4 (Vest and Caldwell, 1972) all show restricted nodulation and non-nodulation with specific strains of *Bradyrhizobium*. In contrast, rj_1 (Williams and Lynch, 1954), and rj_5 (Pracht et al, 1993) soybean lines display a non-nodulation phenotype that is independent of inoculation strain. Naturally occurring non-nodulation mutations also exist in other legumes such as chickpea and pea (Davis et al, 1985; Kneen and LaRue, 1984).

Supernodulating and nonnodulating mutants are important tools for understanding the nodulation processes in the wildtype. The roadblock in these studies is locating these genes for the purposes of sequencing and further determination of their biochemical and physiological function. Once this is surmounted, these genes can then be exploited for

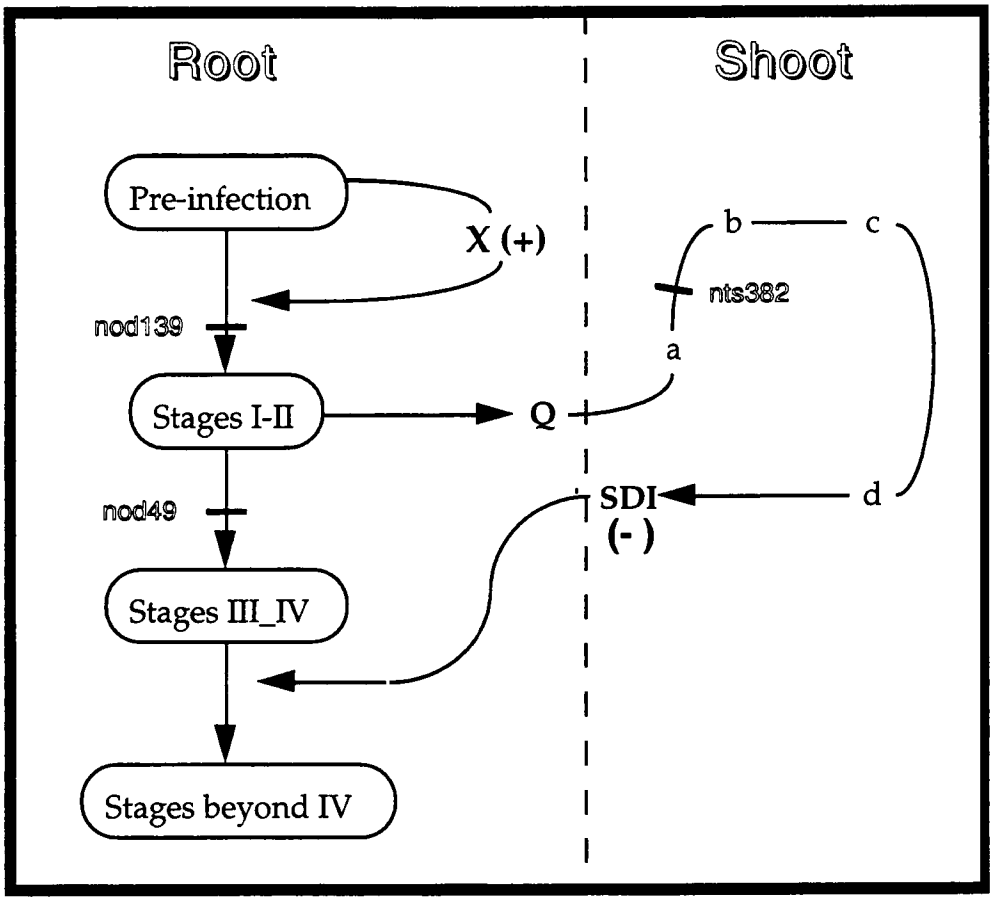


Figure (1-1): Actions of nodulation mutants during the stages of nodulation.

The "X" signal is a positive regulator which stimulates the onset of nodulation and is perhaps related to initiation. "Q" is a translocatable signal to the shoot that nodulation has been initiated. SDI (Shoot Derived Inhibitor) is the autoregulatory signal produced by the shoot to prevent further nodulation events, a-d are putative precursors of SDI. Steps blocked by each mutant are indicated.

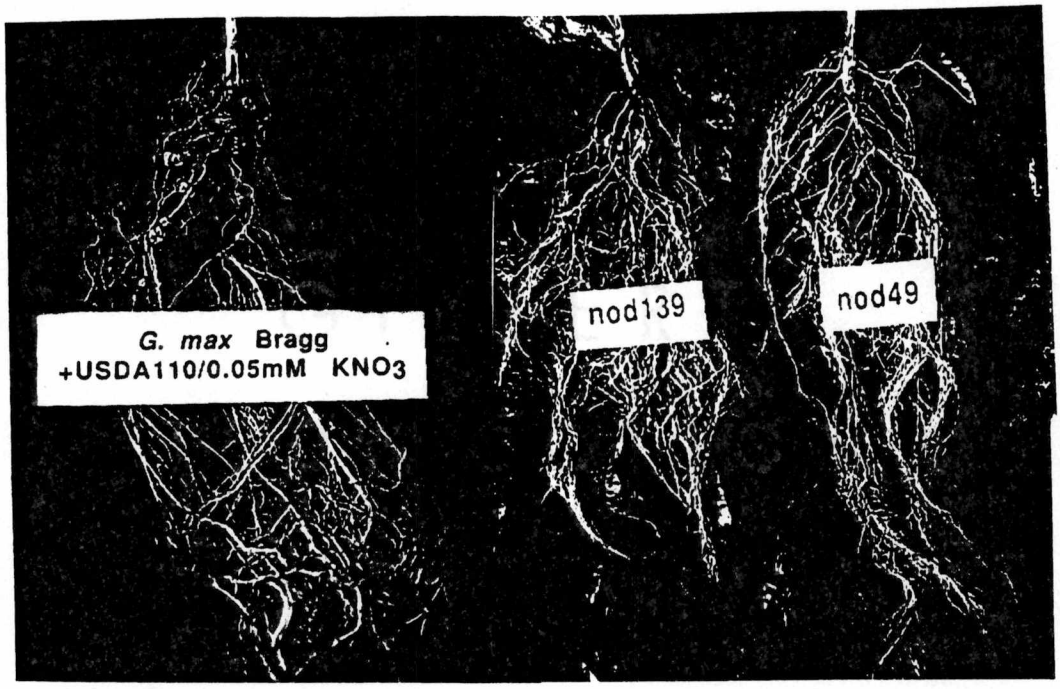


Figure (1-2). Roots of cv. 'Bragg', nod139 and nod49.

All roots have been inoculated and grown in sand-vermiculite with low (0.5 mM) nitrate supplementation. Nod49 shows occasional nodulation while nod139 shows no nodulation

many beneficial purposes. Some future goals are: 1) to produce soybean lines that are specifically infected by more efficient lines of *Bradyrhizobium* (Devine and Breithaupt, 1980), 2) to genetically engineer soybeans which to grow in poorer soils, 3) to create a legume with higher nitrogen fixation capability which can provide neighboring non-legumes with fixed nitrogen (Mulungoy and Akobundo, 1990, Bohlool, 1990), and 4) to produce a tight symbiosis between a genetically engineered *Bradyrhizobium* and a non-nodulating mutant for the purposes of eliminating competitive, but inefficient strains of *Bradyrhizobium*. These all have great potential to make agricultural techniques more economical and environmentally friendly.

II. Non-nodulation mutants

Nod49

Nod49, a mutant of *G. max* (L.) Merr cv 'Bragg', was isolated using ethyl methanesulphonate (EMS) induced mutagenesis and M_2 family selection (Carroll et al, 1986). The non-nodulation phenotype is monogenic, recessive and has been stable for many generations. This phenotype is also independent of the inoculation strain (Carroll et al, 1986). Nod49 lacks root hair curling, and root hair deformation (Mathews et al, 1987), but has been shown to secrete nod-gene inducing root exudates. Before and after inoculation, root exudates and extracts were shown to be similar to that of 'Bragg' (Matthews et al, 1989a). Nod49 does produce some cortical cell divisions, but they do not progress to form nodules (Mathews et al, 1989a). Nod49 was shown to be allelic to the naturally occurring rj_1 (Mathews et al, 1989a), and both rj_1 and nod49 produce occasional nodules, but only at much higher inoculation doses (LaFavre and Eaglesham, 1984, Mathews et al, 1987). In approach graft experiments, nod49 was grafted to cv. 'Bragg'. Results of these experiments showed that with nod49 as the "sensor" root of inoculation, nodulation was suppressed through the systemic autoregulation circuit on the 'Bragg' side

(Caetano-Anollés and Gresshoff, 1990). Both observations suggest that the nod49 mutation blocks nodulation later in the signal reception/transmission process.

A population segregating for the non-nodulation phenotype was produced by crossing nod49 (female) [*G. max* (L.) Merr.] and *G. soja* (male) [PI468.397]. This F_2 population (G43) consists of 27 nod (-) plants and 49 nod (+) plants. F_1 hybrids were confirmed by RFLP analysis, the presence on Nod+ phenotype, and purple flower color (Mathews et al, in review).

Nod139

The second mutant, nod139, also shows monogenic, recessive inheritance of the non-nodulation phenotype, but differs from nod49. Nod139 was induced by EMS mutagenesis (Carroll et al, 1986). It is not allelic to either nod49 or nod772, so it must represent a separate complementation group or gene (Mathews et al, 1989b). No cortical cell divisions occur in nod139, nor is there even occasional nodulation at high inoculation doses (Mathews et al, 1989b). Like nod49, it lacks root hair curling, and root hair deformation, but it does retain nod-gene inducing ability (Mathews et al, 1987, Sutherland et al 1990). Nod139 segregates independently from nod 49 and nts382 (Mathews et al, 1989b).

In approach graft experiments, nod139 was grafted to cv. 'Bragg'. Results of these experiments showed that with nod139 as the "sensor" root of inoculation, nodulation was not suppressed on the 'Bragg' side (Caetano-Anollés and Gresshoff, 1990). This suggests that the nod139 mutation blocks nodulation very early in the signal reception/transmission process. Work by Xie et al, (1995) suggested that nod139 has the ability to recognize the presence of *Bradyrhizobium*. Xie et al studied the stimulation of mycorrhizal colonization of soybeans. When roots of cv. 'Bragg' were co-inoculated with *Bradyrhizobium* and mycorrhizae, increased mycorrhizal colonization was observed. When the experiment was

repeated using nod139, the results were identical. Co-inoculation with a mutated strain of *Bradyrhizobium* that was unable to produce Nod factors, did not show increased colonization. These data can be explained in two ways: 1) nod139 recognizes Nod factors or, 2) the increase in mycorrhizae colonization is caused by a stress response to the *Bradyrhizobium* infection. If the first suggestion is true, *nod139* is not a mutation in a receptor. After applying this information to the model proposed by Stacey et al, (1995), the mutation is most likely in a signal produced after the Nod factors have bound to their receptors which stimulates cortical cell division.

Nod139 [*G. max* (L.) Merr cv 'Bragg'] was crossed with *G. soja* (PI468.397), with nod139 being the female, to create the F_2 population M70. The expected 3:1 ratio of Nod(+) to Nod(-) plants of a single, recessive gene was not seen. The observed 15:1 ratio suggested a diallelic segregation (Mathews et al, in review). When backcrossed to the parent 'Bragg', the ratio was 3:1. NN5, a *N*-nitroso-*N*-methylurea induced, nonnodulating mutant of cultivar 'Williams', was crossed to nod139. The F_1 plants did not nodulate, indicating that NN5 is allelic to nod139 (Pracht et al, 1993). Crosses of NN5 with cv "Harosoy 63" suggested that NN5 carries a deficiency at a second locus called rj_5 (Pracht et al, 1993). There appears to be two loci controlling the non-nod phenotype seen in nod139. 'Bragg' is a spontaneous mutant at one (rj_5) and through mutagenesis, nod139 was mutated at the second locus. The mutated loci in nod139 and NN5 is now called rj_6 (Pracht et al, 1993). From the 15:1 segregation ratio, it appears that *G. soja* is dominant in both rj_5 and rj_6 (Mathews et al, in review, Gresshoff, 1995). This could be a representation of the ancestral tetraploidy in soybean.

III. Genetic markers and their detection

Genetic markers are polymorphic regions of DNA which contain random insertions, deletions, or point mutations. They are used to define and map genomes as well as in genetic diagnostic tests. Genetic markers can be used to locate particular genes for which only the phenotype is known. This strategy is known as positional cloning (Wicking and Williamson, 1991). Once closely linked markers are identified, the genetic and physical distance to the target gene can be established. Then large DNA segments, such as Yeast Artificial Chromosomes (YACs), containing flanking markers, can be cloned (Gresshoff, 1995). *Restriction Fragment Length Polymorphism* (RFLP) (Botstein et al, 1980) was the original method by which polymorphisms were detected. In RFLP, the DNA is cut, separated on agarose gels by electrophoresis, then blotted to a membrane where it is screened with radiolabeled or fluorescent probes. RFLP markers are often co-dominant, so there is no null allele and hybrids can be detected. RFLP is a good technique but problems come from the fact that it uses radioactivity and a large amount of DNA. The amount of data produced is small compared to the amount of work, making it very costly. The use of *microsatellites* (Akkaya et al, 1992) has become popular. Microsatellites are short tandem repeats (STR) in the DNA consisting of 2-7 base pair repetitions which can be detected by Polymerase Chain Reaction (PCR). Their strong point is that most microsatellites have many different alleles, which exist as a different number of STRs. The weakness of microsatellites is that they also require sequence information and investment to be recognized.

Newer methods are employing PCR techniques using arbitrary primers. The advantages to these techniques are simplicity and the fact that they require no prior knowledge of the DNA sequence. *Random Amplified Polymorphic DNA* (RAPD) analysis amplifies nucleic acids with arbitrary primers that are usually 10 nucleotides long. The products are separated

on agarose and stained with ethidium bromide. RAPD primers produce 5 - 10 products per primer, so the amount of information gained is limited (Williams et al, 1990). *Arbitrarily Primed - PCR* (AP-PCR) (Welsh and McClelland, 1990) uses oligonucleotides of length comparable to those used in PCR. This produces 3 - 50 bands which are separated on polyacrylamide gels and visualized by autoradiography. This makes the technique costly and time consuming. *Amplification Fragment Length Polymorphism* (AFLP) (Vos et al, 1995) uses genomic DNA cut with two restriction enzymes, then ligated to adapter sequences. The annealling site for the primer is then the adapter sequence plus 1-3 nucleotides of the restriction site. This allows use of primers that are arbitrary at the 3 prime end, but are also specific for the adapter sequence at the 5 prime end. AFLP generates many bands, which raises the efficiency over that of other methods, but expense and time consumption are still excessive.

The marker detection method employed for this project was *DNA Amplification Fingerprinting* (DAF) (Caetano-Anollés et al, 1991b). DAF is a technique that was developed and patented (US. 5,413,909) here in the Plant Molecular Genetics lab of the University of Tennessee. It also relies on annealing of arbitrary primers to the template during amplification, but it differs in several ways from other arbitrary primer techniques. By 1) increasing the primer to template ratio, 2) shortening the primers and, 3) using polyacrylamide gels and silver staining for visualization. DAF uses primers 5 - 25 nucleotides long (Caetano-Anollés et al, 1992), as well as mini-hairpin primers (Caetano-Anollés and Gresshoff, 1994) which potentially allow the use of a 3 base pair annealing site. A diagram of a mini-hairpin structure is given in Chapter 3. By shortening the primer, more annealing sites are, theoretically, created. An advantage of this is an increase in total number of amplified products. With mini-hairpins, that number can be increased to more

than 75 products per primer. These characteristics allow more of the genome to be scanned per primer, making this method much more efficient in marker detection.

IV. Bulked segregant analysis (BSA)

As genetic maps become saturated, certain genomic regions will become clustered with genetic markers, leaving other regions sparsely covered. A second problem in genome mapping lies in locating genes that segregate in populations for which no map exist (Michelmore et al, 1991). Bulk Segregant Analysis (BSA) was developed as a method for the rapid identification of genetic markers linked any particular region of a genome. The identification of markers linked to a target locus requires the screening of large populations. The number of samples can quickly become too large to work with, but this number can be reduced by using DNA pooling methods. Pools can be based on phenotypes (Michelmore et al, 1991), or previously isolated genetic markers that flank the target region (Giovannoni et al, 1991). BSA allows the screening an entire population with two bulked samples (figure 1-3). The samples are bulked by taking equal amounts of DNA from individuals of a segregating F_2 population and pooling them based on the segregating phenotype. The bulks are then similar to each other at all loci except the loci which control, or are closely linked to the bulking phenotype. All unlinked polymorphic regions are masked by the bulking and the analysis is then focused on the selected locus (Michelmore et al, 1991). In this study, the target loci were *rj1* (nod49) and *rj6* (nod139) and the phenotype was nodulation ability. Once distinguishing markers are established in the bulks, the individuals can be scored to determine the linkage distance.

In pooling strategies, markers can be detected within a certain interval, called the "window" of linkage detection (Williams et al, 1990). The size of the window is determined by two characteristics: 1) bulk size, and 2) band amplification potential. Bulk size and window size

DAF Amplification Profiles

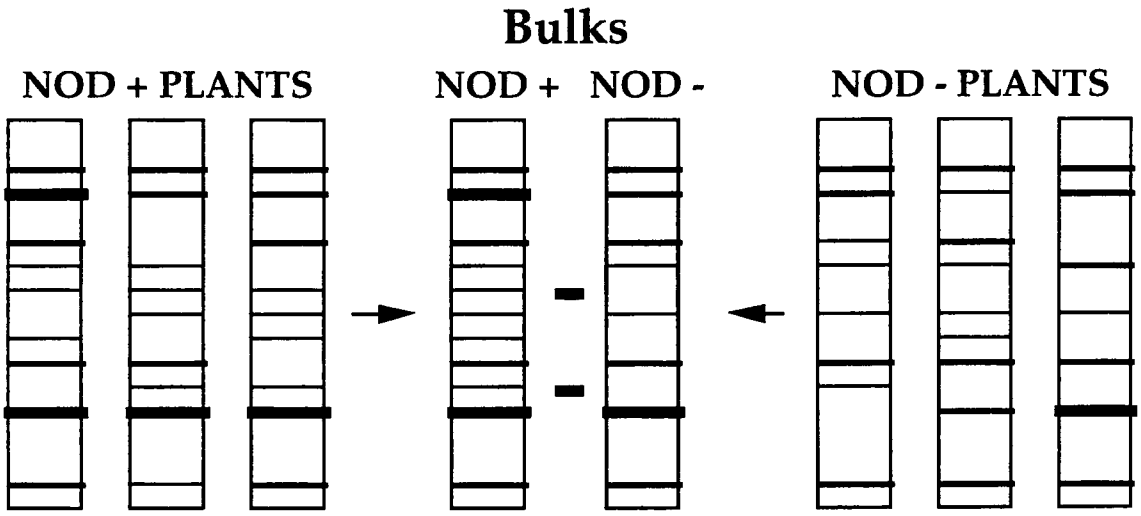


Figure (1-3): Strategy involved in Bulked Segregant Analysis.

BSA works by isolating genomic regions linked to a phenotype in one bulk. This pooling of the DNA from individual plants allows only loci linked to a phenotypic trait to be detected as polymorphic, unlinked loci will appear in both bulks. BSA increases efficiency of template screening by creating a situation where an entire population can be screened using two samples.

are inversely related. Large bulks have a small window and small bulks have a large window. With large bulks, only very closely linked markers will be detected (Tinker et al, 1994). Second, the ease of amplification for a band is called the "amplification potential" (Tinker et al, 1994). When amplification potential is low, a marker that is rare in a pool will not be detected. If the amplification potential for that marker is high, it would be detected even though it existed in a very small portion of the bulk. High amplification potential can mask a legitimate marker from detection by allowing a band to amplify at a low segregation percentage.

BSA was first used to identify RAPD markers for downy mildew resistance in lettuce (Michelmore et al, 1991). Since then, this technique has been used in many different systems: 1) for sex determination markers, *Pistacia vera* (Hormaza et al, 1994), 2) powdery mildew resistance in pea (Timmerman et al, 1994), 3) gall midge resistance in rice (Mohan et al, 1994), 4) tomato ripening markers (Gionvannoni et al, 1991) and, 5) resistance markers in barley (Barua et al, 1993). On average, these studies found one linked RAPD marker for every 100 primers screened. Since the RAPD technique produces so few products per primer, it was hypothesized that by using DAF, fewer primers would be needed to find a linked marker. This hypothesis is based on knowledge that DAF is a much more efficient detection method than RAPD, because of the number of products, and the large number of DAF polymorphisms that exist between *G. max* and *G. soja* (Prabhu and Gresshoff, 1994).

V. Project goal

This study was aimed at isolating genetic markers linked to the loci controlling nod49 and nod139 by coupling BSA with DAF. This was the first project to use these two methods in combination. The assumption was that by combining the quick isolation of BSA and the

efficient detection associated with DAF, many linked markers could be found in a small amount of time. After exhausting the supply of 319 mini-hairpin primers in a bulk screen, it was assumed that problems existed with this combination, possibly resulting from the nature of DAF markers or conservation in this region of the genome. The focus of the project then turned toward finding explanations for why linked markers could not be isolated. Through competition experiments and further analysis of these techniques, possible explanations have been found. This information has value to future projects in order to adjust the technique the needs of individual projects. This may involve the modification of bulk size in BSA, or stringency in DAF.

Chapter 2

Materials and Methods

I. Plant tissue collection.

nod 49

F_2 plants were grown in the Racheff greenhouse in sand with a vermiculite topping and inoculated with *Bradyrhizobium japonicum* strain USDA110. After the plants had grown for four weeks, they were removed from the sand-vermiculite mixture and screened for the presence of nodules. This was performed to separate the nod- (rj_1rj_1) plants from the nod+ plants for bulking purposes. A list of the genotypes of the different generations of the G43 population is provided in table (2-1). They were then replanted in Fafard #4 soil mix and allowed to mature for seed collection. Young leaves were then picked from each plant for DNA isolation. Younger leaves contain fewer secondary products than mature leaves which reduce the quality of the extracted DNA.

nod 139

F_2 plants were grown and screened in the same way as the nod49 plants. A minimum of ten F_3 progeny of 51 nodulating F_2 plants were grown and screened for segregation of the nodulation phenotype. F_2 nodulated plants whose F_3 progeny did not show segregation of the nodulation phenotype were used in the nod+ bulk to reduce the amount of heterozygous material. F_2 plants having progeny that did not nodulate were assumed to be homozygous recessives and used for the nod- bulk. A list of M70 progeny genotypes is provided in table (2-2). The non-segregating F_3 plants were grown to maturity and young leaves were taken for DNA isolation.

P1 generation: G. soja (PI468.397) x nod49 (Bragg)

Rj_1Rj_1 x rj_1rj_1

F_1 generation: Rj_1rj_1

F_2 generation:

	Rj_1	rj_1
Rj_1	**Rj_1Rj_1**	**Rj_1rj_1**
rj_1	**Rj_1rj_1**	rj_1rj_1

Table (2-1): G43 generation genotypes

Genotypes shown in bold were included in the nod+ bulk. Both heterozygous and homozygous material was used in the nod+ bulk. The nod- bulk consisted of only homozygous material.

P1 generation: G. soja (PI468.397) x nod139 (Bragg)

Rj5Rj5Rj6Rj6 x *rj5rj5rj6rj6*

F_1 generation: *Rj5rj5Rj76rj6*

F_2 generation:

	Rj5Rj6	*Rj5rj6*	*rj5Rj6*	*rj5rj6*
Rj5Rj6	***Rj5Rj5Rj6Rj6***	***Rj5Rj5Rj6Rj6***	***Rj5rj5Rj6Rj6***	*Rj5rj5Rj76rj6*
Rj5rj6	***Rj5Rj5Rj6rj6***	***Rj5Rj5rj6rj6***	*Rj5rj5Rj76rj6*	*Rj5rj5rj6rj6*
rj5Rj6	***Rj5rj5Rj6Rj6***	*Rj5rj5Rj76rj6*	***rj5rj5Rj6Rj6***	*rj5rj5Rj76rj6*
rj5rj6	*Rj5rj5Rj76rj6*	*Rj5rj5rj6rj6*	*rj5rj5Rj76rj6*	*rj5rj5rj6rj6*

Table (2-2): M70 generation genotypes

Genotypes shown in bold face will not show segregation of the nodulation phenotype in further generations and cannot be screened out. Therefore the nod+ bulk can never be totally homozygous. The bold face genotypes were also the plants included in the nod+ bulk. The nod- bulk included only the homozygous recessive material that did not nodulate in the F_2 generation.

II. DNA extraction and isolation

Chromosomal DNA was isolated according to Dellaporta et al., (1983). Approximately 0.5 g of young leaves were picked and frozen in liquid nitrogen. These leaves were then ground in liquid nitrogen. Polyclar AT (0.3 to 0.5 g) was added to the powdered leaf and liquid mixture. This mixture was placed in a 50 mL plastic centrifuge tube and 15 mL of extraction buffer (0.1 M Tris-HCl pH 8.0, 50 mM EDTA, 0.5 M NaCl) was added. One mL of 20% SDS was then added and the solution was mixed and placed in a 65° C water bath for one hour. Five mL of 5 M potassium acetate (pH 8.0) was then added and the solution was mixed. The solution was placed on ice for 30 minutes to allow the SDS and potassium complexes to precipitate. The mixture was then poured through pleated filter paper into clean 50 mL centrifuge tubes. Ten mL of cold propanol (stored in -20° C) was then added and the mixture was then put into the -20° C freezer for 30 minutes. The tube was then centrifuged for 15 minutes at 4° C using the JA20 rotor at 11,000 rpm. The supernatant was poured off and the tube was left in the fume hood with the lid off to allow the propanol to evaporate. When the odor of propanol was gone from the tube, 0.7 mL of chromosomal DNA buffer (50 mM Tris-HCl pH 8.0, 10 mM EDTA) was added to the pellet. This was then put on ice until the pellet desolved. The solution was transferred to a 1.5 mL Eppendorf tube and centrifuged for 10 min. at 12,000 rpm to precipitate polysaccharides and SDS complexes. The supernatant was transferred to a new 1.5 mL Eppendorf tube and 75 µL of 3 M sodium acetate (pH 7.5) was added. Then, 0.5 mL of propanol was added and the tube was mixed gently. The tube was centrifuged at full speed for 2 minutes and the supernatant poured off. The pellet was washed twice with 70 - 80% cold (4° C) ethanol to remove salts. The tube was allowed to stand 5 minutes before being centrifuged at full speed for 2 minutes. The ethanol was

poured off and the pellet was dried in the Speed Vac concentrator. The pellet was then dissolved in 100 μL of sterile dH_20 and stored at 4° C.

III. DNA quantification

The amount of isolated DNA was quantified using a micro-assay performed with a dedicated TKO100 mini-flourometer (Hoefer Scientific Instruments, San Francisco, CA). The Fluorometer measures the amount of H33258 dye that is taken up by double stranded DNA. The working dye was made by adding 10 μL of dye stock (1mg/mL in H_20) to 100 mL of 1X TNE buffer (0.1 M Tris, 10 mM EDTA, 1.0 M NaCl, pH 7.4) to make a final concentration of 0.1 μg/mL. The assay was then performed by adding 2 μL of sample to 2 mL of working dye. The digital display then indicated the DNA concentration in ng/μL. The samples were assayed only after the fluorometer had warmed up for 30 minutes and was standardized with reference samples of calf thymus DNA in the expected concentration ranges of the sample DNA.

IV. Bulking of DNA samples

DNA samples were bulked based on the nodulation phenotype. Those nod49 samples that had produced nodules were placed in the N+ bulk (38 plants), and samples that had produced no nodules were placed in the N- bulk (26 plants). Nod139 bulks were created in the same fashion except, only ten plants were used in each of the bulks which were termed A+ and A-. There were concerns about the original bulking of nod49. It had been done by taking 200 ng of DNA from each of the concentrated samples. The problem in this was that by taking the aliquots from such concentrated material, accurate pippetting was improbable. After this realization, the samples of both nod49 and nod139 were diluted to approximately 100 ng/μL and then quantified again. They were then diluted to

approximately 20 ng/μL and quantified a third time. From there, 60 ng from each sample was placed into the appropriate bulk. The bulks were then diluted to a working concentration of 2 ng/μL.

V. DNA amplification fingerprinting (DAF)

The amplifications were carried out in a volume of 20 μL which contained 4 ng of DNA (2μL of 2 ng/μL stock, 10.8 μL of ddH_2O, 3 μM primer (2 μL of 30 μM stock), 0.4 μL Amplitaq Stoffel fragment DNA polymerase (PE-Cetus, Norwalk, CT), 200 μM of each deoxynucleoside triphosphates (2 μL of dNTP stock), 4 mM $MgSO_4$ (0.8 μl of 0.1 M stock), and 2 μL TTNK buffer [10 mM KCl, 4 mM $(NH_4)_2SO_4$, 0.1 % Triton X-100, and 20 mM Tris-HCl (pH 8.3)] (Caetano-Anollés and Gresshoff, 1994). The amplifications were performed using the mini-hairpin primers in table (2-3) (Caetano-Anollés and Gresshoff, 1994). The primers contained a hairpin sequence and a three nucleotide core (see Chapter 3 for structure). Four sets of sixty-four had the same hairpin sequence with the exception of a one base substitution in the loop of the hairpin. The fifth set of primers had a shorter hairpin sequence (see Chapter 3 for structure). The reaction was overlaid with one drop of mineral oil (Mallinckrodt). The amplification was carried out in an Ericomp twin block thermocycler from the Easycycler series (San Diego, CA). The profile used was 35 cycles of 95° C for 30 sec., 30° C for 30 sec., 72° C for 30 sec., then one cycle of 72° C for 5 minutes (fig 2-1). The amplified products were recovered from the oil by chloroform extraction. Chloroform (100 μL) was added to each tube containing the amplification mix and oil. A bubble, containing the amplification mix, was extracted and placed in a new Eppendorf tube and stored at 4°C.

Table (2-3): Primer sequences.

HpA hairpin	HpB hairpin	HpC hairpin	HpD hairpin
GCGAAAGC-	GCGACAGC-	GCGAGAGC-	GCGATAGC

HP hairpin
GCGAAGC-

Core #	Core	Core #	Core	*Core #*	*Core*	*Core #*	*Core*
1	-CCC	33	-GCC	*8*	*-GCT*	*58*	*-TGA*
2	-CCT	34	-GCT	*10*	*-CTG*	*59*	*-TAC*
3	-CCG	35	-GCG	*28*	*-CTA*	*60*	*-TAT*
4	-CCA	36	-GCA	*29*	*-CTC*	*61*	*-TAG*
5	-CTC	37	-GTC	*30*	*-CTT*	*62*	*-TAA*
6	-CTT	38	-GTT	*31*	*-CAG*	*63*	*-GCC*
7	-CTG	39	-GTG	*32*	*-CCG*	*64*	*-GCG*
8	-CTA	40	-GTA	*33*	*-CGG*	*65*	*-GCA*
9	-CGC	41	-GGC	*34*	*-CAC*	*66*	*-GTC*
10	-CGT	42	-GGT	*35*	*-CGC*	*67*	*-GTT*
11	-CGG	43	-GGG	*36*	*-CCC*	*68*	*-GTA*
12	-CGA	44	-GGA	*37*	*-CGT*	*69*	*-GGC*
13	-CAC	45	-GAC	*38*	*-CCT*	*70*	*-GGT*
14	-CAT	46	-GAT	*39*	*-CAT*	*71*	*-GGA*
15	-CAG	47	-GAG	*40*	*-CGA*	*72*	*-GAC*
16	-CAA	48	-GAA	*41*	*-CCA*	*73*	*-GAT*
17	-TCC	49	-ACC	*42*	*-CAA*	*74*	*-GAG*
18	-TCT	50	-ACT	*43*	*-GTG*	*75*	*-GAA*
19	-TCG	51	-ACG	*44*	*-ATG*	*76*	*-ACC*
20	-TCA	52	-ACA	*45*	*-TTG*	*77*	*-ACT*
21	-TTC	53	-ATC	*46*	*-GGG*	*78*	*-ACG*
22	-TTT	54	-ATT	*47*	*-AGG*	*79*	*-ACA*
23	-TTG	55	-ATG	*48*	*-TGG*	*80*	*-ATC*
24	-TTA	56	-ATA	*49*	*-TCC*	*81*	*-ATT*
25	-TGC	57	-AGC	*50*	*-TCT*	*82*	*-ATA*
26	-TGT	58	-AGT	*51*	*-TCG*	*83*	*-AGC*
27	-TGG	59	-AGG	*52*	*-TCA*	*84*	*-AGT*
28	-TGA	60	-AGA	*53*	*-TTC*	*85*	*-AGA*
29	-TAC	61	-AAC	*54*	*-TTT*	*86*	*-AAC*
30	-TAT	62	-AAT	*55*	*-TTA*	*87*	*-AAT*
31	-TAG	63	-AAG	*56*	*-TGC*	*88*	*-AAG*
32	-TAA	64	-AAA	*57*	*-TGT*	*89*	*-AAA*

Primer sequence is found by taking the series core sequence and adding the numbered sequence (ex. HP63 is GCGAAGC-GCC). HpA, HpB, HpC, and HpD all share the same tag numbering system. The HP series tag numbering system is in italics.

VI. DNA separation and silver staining

Separation of the amplified products was done by electrophoresis on a 10% polyacrylamide gel backed with Gelbond PAG polyester film (FMC, Rockland, Maine). The electrophoresis was carried out in a Bio-Rad Mini-Protean II apparatus. The formula for a stock gel solution (400 mL) was: 39.2 g acrylamide, 40 g urea, 20 mL glycerol, 40 mL 10X TBE buffer (121.1 g Tris base, 51.35 g boric acid, and 3.72 g EDTA in 1L of dH_2O, pH 8.3), 0.8 g PDA, and dH_20 to 400 mL (Bassam and Bentley, 1995). The formula for the loading buffer (100 mL) was: 40 g urea, 3 g ficoll, 0.12 g Tris base, 1.12 g EDTA, 0.02 g Xylene Cyanol, 0.02 g Bromophenol Blue, and dH_20 to 100 mL (Bassam and Bentley, 1995). Ten µL of the gel solution was polymerized with 15 µL of TEMED and 150 µL of 10% Ammonium persulfate. This was enough to cast two gels. Amplified samples, and the reference Biomarker, were individually mixed with the loading buffer at a ratio of 3 µL of sample : 3 µL of loading buffer. The reference marker contained bands at 1000, 700, 525, 500, 400, 300, 200, 100, and 50 bp. The reference was included in the outside lanes of most gels. Once loaded, the gels were run at ~150V for 1 hr or until the second dye front was even with the electrode in 1X TBE buffer (~1 cm from bottom).

Visualization of the DNA was obtained by silver staining. Because the DNA is negatively charged, the positive charge on the silver ion allows it to bind to the DNA. The silver is then reduced by formaldehyde in a sodium carbonate solution until bands becomes visible. This reduction is then stopped by adding acetic acid. Silver staining (Bassam et al, 1991) was performed the following way. The gels were covered and shaken in 7.5 % acetic acid for 10 minutes, and washed 3 times for 2 minutes in dH_20. The gels were then stained with a silver solution containing 0.1 g/100 mL of silver nitrate and 150

µL/100 mL of formaldehyde for 20 minutes. The gels were then washed for 20 seconds in dH_20 and developed in a cold (~8° C) solution containing 3 g sodium carbonate/ 100 mL, 50 µL/100 mL of sodium thiosulfate stock (0.2 g/50 mL), and 300 µL/100 mL of formaldehyde. The reaction was stopped by immersing the gels in 10% acetic acid for 1 - 2 minutes and then washing with dH_20 for five minutes. The gels were air dried overnight for preservation (Bassam et al, 1991).

During the project advances were made in the staining protocol. After the three dH_2O washes, the gels were soaked in a 18.5 % formaldehyde solution for ten minutes followed by a five minute dH_2O wash. Then stained with a solution only $AgNO_3$ (0.1 g / 100 µL), then developed with a $NaCO_3$ solution (3g / 100 µL) that was not chilled. These changes prevented the edges of the gel from staining and allowed the use of simplified solutions.

Gel cracking was experienced during winter due to a humidity drop in the laboratory. This was battled by using an "anti-cracking" solution (1 % glycerol, 35 % ethanol, and 10 % acetic acid) (Dr. Natasha Kruchinina, pers. comm.). This solution was placed on the gels after the final wash of the staining for 5 minutes. The gels were then hung to dry. This solution allowed the gels to retain enough moisture to keep them from cracking and curling.

VII. Scoring of DAF products.

Bands were scored visually by counting bands that were present in one bulk and absent in the other. Only products between 700 bp and 100 bp were scored because of a lack of reliability in bands outside that range. In the case of nod49, bands were scored as polymorphic only if they came from the nod+ bulk. This was due to the heterozygous

material included in the bulks (see table 2-1). For nod139, bands coming from either bulk were scored as polymorphic due to a more homozygous nature to the nod139 bulks (see table (2-2) for explanation of homozygous nature).

Chapter 3

Results

I. Progeny screen of nod139

The F_2 generation produced 170 nodulating plants and 17 non-nodulating plants. Chi-square analysis reveals that this was not statistically different from the expected 15:1 segregation ratio. Results of a screen of approximately 700 F_3 plants show that of the original 51 nod^+ F_2 plants, 19 showed no segregation of the nod^+ phenotype and 23 showed segregation of the nodulation phenotype (nod^+ and nod^-). Inconclusive data were obtained for 9 plants which showed a reduced number or atypical nodulation characteristics, such as one or two large nodules. Ten of these 19, presumed to be homozygous dominant, were used in the nod+ bulk for screening. The individuals presumed to be homozygous recessive from the F_2 screen, were used in the nod- bulk.

II. Parental screening

DNA Polymorphisms between *G. soja* and nod49 were screened using a randomly chosen subset of 27 mini-hairpin primers. Those 27 primers produced 48 polymorphic bands, which gave an average of 1.7 polymorphisms per primer. This compares well with the findings of Prabhu and Gresshoff (1994) who found 40 polymorphic bands with 26 octamer primers for an average of 1.56 bands per primer when comparing cv. Bragg to *G. soja* (PI468.397), as done in this study. Further results are listed in table (3-1) a summary of that table is listed below in table (3-2).

Table (3-1): Primer screen of parental lines of *G. soja* (PI468.397) and nod49 (Bragg).

Primer	Total # of bands	# of polymorphic bands
Hp A5	37	2
Hp A10	33	1
Hp A19	32	0
Hp A28	41	3
Hp A37	36	2
Hp A40	40	2
Hp A48	39	2
Hp A54	42	1
Hp A47	44	2
Hp B12	39	1
Hp B19	45	2
Hp B24	45	2
Hp B30	44	2
Hp B37	50	5
Hp B38	39	0
Hp B42	39	2
Hp B48	45	1
Hp B62	40	2
Hp C8	42	1
Hp C19	46	2
Hp C38	45	2
Hp C39	43	3
Hp D6	54	2
Hp D7	42	2
Hp D9	43	1
Hp D23	46	2
Hp D49	49	1

Summary:

Total # of bands	1140
# of polymorphic bands	48
% of polymorphic bands	4.2
# of polymorphisms/ primer	1.7
# of polymorphism/ 319 primers	542.3

Primers were selected randomly.

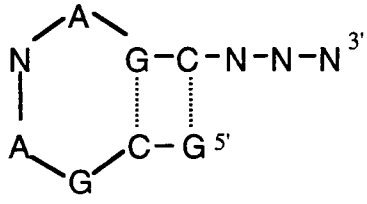
Table (3-2): Summary of mini-hairpin parental screen data.

Total # of bands	1140
# of polymorphic bands	48
% of polymorphic bands	4.2
# of polymorphisms/ primer	1.7
# of polymorphism/ 319 primers	542

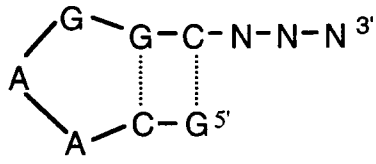
III. Mini-hairpin screening

The nodulation bulks of both nod49 and nod139 were screened with a library of 319 mini-hairpin primers. 255 of these mini-hairpins have structure (A), the other set of 64 have structure (B).

A.



B.



All possible combinations of nucleotides at the "N" locations in structure A render 256 different primers. In structure B, 64 different primers can be generated. The annealing site is then assumed to be the three base pair core at the 3' end. Apparently, under these conditions, these sequences are able to form a very stable mini-hairpin structure (Hirao et al, 1992). This occurs by the first two nucleotides from the 5 prime end, GC, annealing to the 7th (G) and 8th (C) nucleotides and form the stem of the loop. This was confirmed in two ways: 1) 20% PAGE gels show that the hairpin has an unusually rapid mobility (Hirao et al, 1992), and 2) using the known sequence of the plasmid, pUC18, predicted amplification products of the 3 nucleotide annealing site were satisfied (Caetano-Anollés

and Gresshoff, 1994). Other evidence that the hairpin-turn structure actually forms comes from amplifications performed with primers containing only the first eight nucleotides from the 5' end. When these primers where used in DAF reactions, no amplification products were detected, although there is ample evidence that 8 nucleotide long primers (without GC stem sites) will produce many products (Prabhu and Gresshoff, 1994). Normal primers of three nucleotides will not amplify (Caetano-Anollés et al, 1992), but through the use of these mini-hairpins, a very small annealing site can be created.

General screening results of the nod49 and nod139 bulks are given in table (3-3). These data indicate that the polymorphic band percentage is low (avg.= 0.22%). Chi-square analysis reveals that the total polymorphic band percentage between nod49 and nod139 is not statistically different because $X^2_c = 0.510 < X^2_t = 3.841$. Chi-square values between individual primer sets are ($X^2_t = 3.841$): HP = 0.4, Hp A = 0.11, Hp B = 1, Hp C = 1.8, and Hp D = 4. Only the Hp D series exceeds the tabular chi-square value of 3.841. This is mostly likely the result of degraded primers. The Hp D series displayed several instances of incomplete amplification which could result in a false detection of a polymorphism.

IV. Bulk screening

nod49

Seventy-five plants of a subset of the nod49 population (G43), where nod49 was used as the female parent, were grown for DNA isolation. The segregation data for the nodulation phenotype were: 48 nod+ plants : 27 nod- plants. Because the nod- plants used for this study were actually a collection from several different growings of G43, the original chi-square data is still valid here. Since the calculated X^2 value (0.0) is less than the tabular X^2 value (3.841), the G43 segregation data is not significantly different from the expected 3:1 confirming the non-nodulation segregation as a Mendelian character. DAF profiles of

Table (3-3): Primer screening of nod49 and nod139 bulks.

A. nod49 derived bulks:

	HP	Hp A	Hp B	Hp C	Hp D	Total
Total # of Bands	1927	2178	2264	2240	2253	10,862
Avg. # of Bands per Primer	30.1	34.0	35.4	35.0	35.8	34.05
Polymorphic Band %	0.31	0.23	0.13	0.18	0.18	0.20
Avg. # of polymorphisms per primer	0.09	0.08	0.05	0.06	0.06	0.07

B. nod139 derived bulks:

	HP	Hp A	Hp B	Hp C	Hp D	Total
Total # of Bands	1802	2509	2237	2398	2405	11,351
Avg. # of Bands per Primer	28.1	39.2	34.9	37.5	38.2	35.6
Polymorphic Band %	0.22	0.16	0.27	0.04	0.50	0.24
Avg. # of polymorphisms per primer	0.06	0.06	0.09	0.02	0.19	0.08

Hp refers to mini-hairpin. There are five mini-hairpin series indicated which have a total of 319 primers. Table A and B show the screening results of the nod49 and nod139 bulks with the five mini-hairpin sets.

genomic DNA from the Nod+ and Nod- pools of the G43 cross were obtained with mini-hairpin primers. Bands were only scored as polymorphic if they were clearly present in the Nod+ bulk and absent in the Nod- bulk. Because the Nod+ bulk contains heterozygous material, bands present in the Nod- bulk and absent in the Nod+ bulk were not considered polymorphic even though they may be linked in repulsion.

Early in the screening, primer HpA47 gave a reproducible polymorphism at 400 bp. In three separate amplifications, the marker was shown to be polymorphic (figure 3-1). The third amplification shows a band present in the Nod- bulk near 400 bp, but closer study revealed that it was smaller than the suspected marker. However, the observation is of little value, as further analysis demonstrated that the 400 bp band existed in both the nod49 and the *G. soja* parent. Segregation analysis of the individuals of nod49 bulks showed 46% recombination (independent assortment is 50%).

Further screening revealed 19 other polymorphic bands. For reasons of reliability, only band differences that were also polymorphic in the parents, nod49 and *G. soja*, were considered to be real candidate markers. Only bands produced by Hp D37 and Hp D50 were also present as polymorphisms in the parents (figure 3-1). Hp D37 produced two polymorphisms, one at 180 bp from the *G. soja* parent, which showed 45% recombination and a second at 175 bp from the nod49 parent which showed 44% recombination. The segregation data for these two bands, suggest that they are most likely not co-dominant. Hp D50 also produced a polymorphic band at 400 bp from the *G. soja* parent which showed 39% recombination. The polymorphic band data are given in table (3-4). The results of the screen of the nod49 bulks and the parental lines are given in table (3-5). Results of the segregation of these polymorphic bands in the individuals is given in table (3-6).

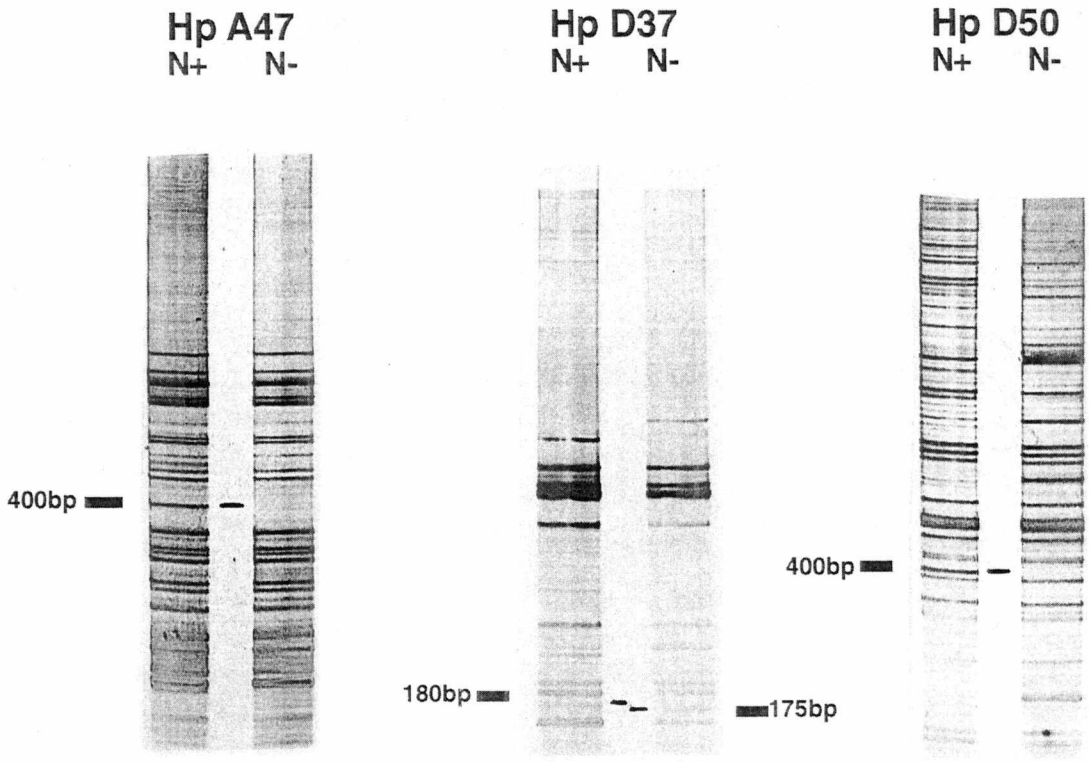


Figure (3-1): Nod49 polymorphic band profiles

These are the profiles of the potential markers from the nod49 bulks that were also present in the parental lines. Molecular weights are not standardized between each profile because each was from a separate gel run. Only true polymorphisms that passed the two screens are indicated. Others that are present may be artifacts or the result of incomplete amplification.

Table (3-4): nod49 polymorphic band data.

Primer	# of (+) plants w/band	# of (-) plants w/band	recombination % (A)	recombination % (B)
Hp A47	7/30	2/26	45 (25/56)	8 (2/26)
Hp D37 (180 bp)	14/38	4/23	45 (29/64)	17 (4/23)
Hp D37 (175 bp)	13/38	11/26	44 (36/64)	42 (11/26)
Hp D 50	31/38	18/26	39 (25/64)	69 (18/26)

Recombination percentage A is based on information from both the nod+ and nod- bulks. Recombination percentage B is based on information from only the nod- bulk. Percentages are calculated by adding the number of segregants in each bulk by the total number of plants.

Table (3-5): Polymorphic band distribution from nod49 bulks and parents.

Primer	Band Size (bp)	Presence in *G. soja*	Presence in nod49	Presence in N+ bulk	Presence in N- bulk
HpA 5	180	Yes	Yes	Yes	No
HpA 10	290	Yes	Yes	Yes	No
HpA 19	230	No	No	Yes	No
HpA 28	310	Yes	Yes	Yes	No
HpA 47	400	Yes	Yes	Yes	No
HpB 42	380	Yes	Yes	Yes	No
HpB 48	280 & 450	Yes, Yes	Yes, Yes	Yes, Yes	No, No
HpC 8	300	No	No	Yes	No
HpC 10	250	Yes	Yes	Yes	No
HpC 19	350	Yes	Yes	Yes	No
HpC 38	400	Yes	Yes	Yes	No
HpD 6	700	No	No	Yes	No
HpD 37	***180,175***	***Yes, No***	***No, Yes***	***Yes, Yes***	***No, No***
HpD 50	***400***	***Yes***	***No***	***Yes***	***No***
Hp 69	380	Yes	Yes	Yes	No
Hp 42	650	Yes	Yes	Yes	No
Hp 43	230	Yes	Yes	Yes	No
Hp 49	310	Yes	Yes	Yes	No
Hp 73	700	Yes	Yes	Yes	No
Hp 80	180	Yes	Yes	Yes	No

Bands that proved to be polymorphic in both the bulks and the parents are in bold face and italicized. Only bands produced in the nod+ bulk and not in the nod- bulk were considered polymorphic because the nod+ bulk contained heterozygous material.

Plant #	Nod (+/-)	HpA47 (400bp)	HpD37 (180bp)	HpD37 (175bp)	Plant #	Nod (+/-)	HpA47 (400bp)	HpD37 (180bp)	HpD37 (175bp)
121	-	-	?	?	107	+	-	?	?
122	-	-	?	?	109	+	-	?	?
123	-	-	-	-	110	+	-	?	?
124	-	-	?	?	111	+	-	?	?
125	-	+	-	-	114	+	-	?	?
126	-	-	-	-	115	+	-	?	?
127	-	-	-	-	116	+	-	?	?
129	-	-	-	-	118	+	-	?	?
131	-	-	-	+	120	+	-	?	?
132	-	-	-	+	149	+	-	?	?
133	-	-	-	-	151	+	-	?	?
134	-	-	-	-	152	+	-	?	?
135	-	+	+	-	153	+	-	?	?
136	-	-	-	-	154	+	+	+	-
137	-	-	-	+	155	+	+	+	-
138	-	-	-	+	156	+	-	-	-
139	-	-	-	-	157	+	-	-	+
140	-	-	-	+	158	+	-	-	+
141	-	-	-	-	159	+	-	-	-
142	-	-	-	-	160	+	-	-	-
143	-	-	+	+	161	+	+	+	-
144	-	-	-	-	162	+	+	+	-
145	-	-	+	-	163	+	-	-	+
146	-	-	-	+	164	+	+	+	-
147	-	-	+	-	168	+	-	-	-
148	-	-	-	-	169	+	+	+	+
100	+	-	?	?	170	+	+	+	-
101	+	-	?	?	172	+	?	?	-
104	+	-	?	?	173	+	?	?	-

Table (3-6): Nod49 individual screen results.

(+) indicates nodulation in Nod(+/-) column and presence of band in all other columns. (-) indicates no nodulation in Nod(+/-) column and absence of band in all other columns. (?) denotes missing data due to no amplification or loss of gel, from cracking, at time of table preparation.

Recombination frequencies were also calculated which were based only on the nod- bulk data. This was done to eliminate the effect of the heterozygous material in the nod+ bulk. For the majority of the polymorphic bands, this improved the linkage greatly. This suggests that these polymorphisms are actually good markers.

nod139

For the nod139 analysis, 187 plants of the M70 population, where nod 139 was also used as the female parent, were grown. The segregation data were: 170 nod+ : 17 nod-. Since the calculated X^2 value (2.57) is less than the tabular X^2 value (3.841), the M70 segregation data is not significantly different from 15:1 indicating Mendelian segregation. DAF profiles were obtained of the nod139 bulks in the same way as with nod49 with the exception of bulk size. Theories of how bulk size affected BSA led to the use of only ten plants in each bulk for nod139. Because of the F_3 screen, it was possible to exclude some heterozygous material from the nod+ bulks.

The nod139 bulks were screened with the same library of mini-hairpin primers as the nod49 bulks. Twenty-seven polymorphic bands were detected in the bulks. Again, for reasons of reliability, only bands that were also polymorphic in the parents, nod139 and *G. soja*, were considered to be real polymorphisms. Only 4 of the 27 also produced polymorphic bands in the parents: Hp C39 (600 bp from *G. soja*), Hp D9 (200 bp from nod139), Hp D23 (200 bp, from nod139), and Hp D39 (220 bp from *G. soja*). They are shown in figure (3-2). The Hp D23 and Hp D9 bands were unique in that they were present in the nod- bulk and not in the nod+ bulk, all others were the reverse condition. The polymorphic band data are given in table (3-7). The nod139 bulk and parental screen results are given in table (3-8). Screening of the individuals with these primers showed that none were closely linked to *nod139* (table 3-9).

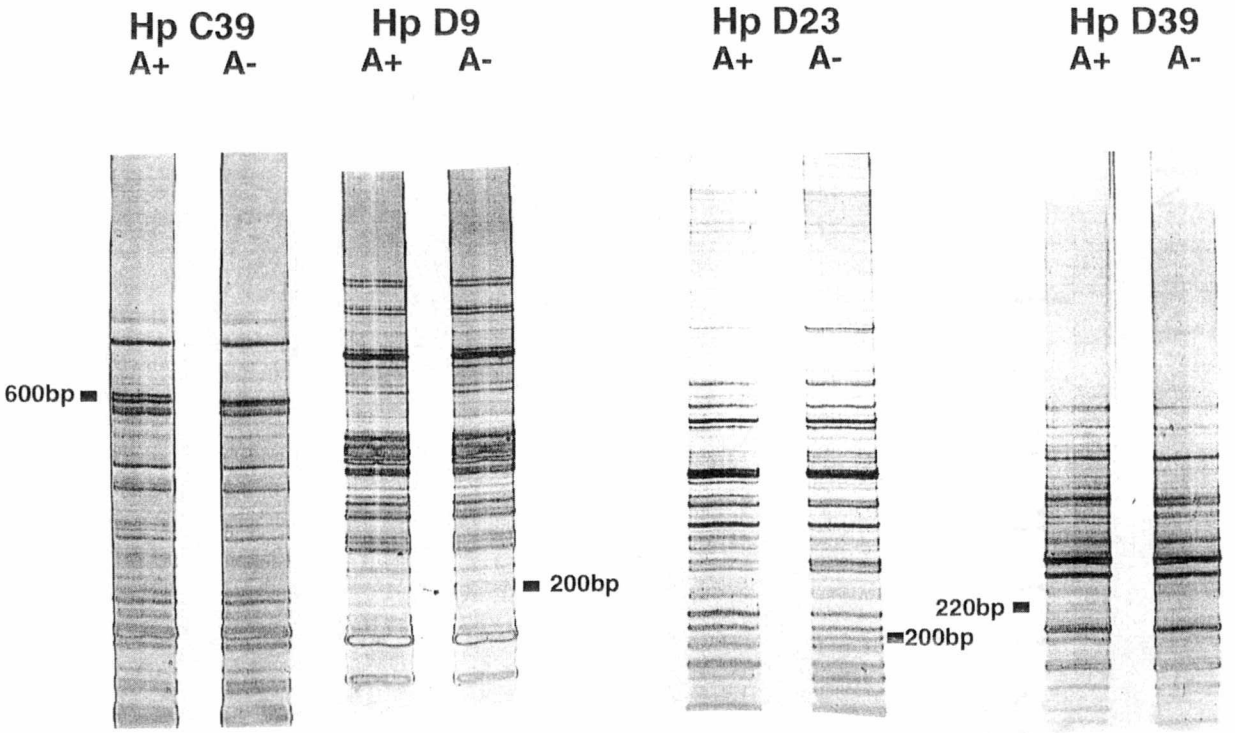


Figure (3-2): Nod139 polymorphic band profiles

These are the profiles of the potential markers from the nod49 bulks that were also present in the parental lines. Molecular weights are not standardized between each profile because each was from a separate gel run. Only true polymorphisms that passed the two screens are indicated. Others that are present may be artifacts or the result of incomplete amplification.

Table (3-7): nod139 polymorphic band data.

Primer	# of (+) plants w/band	# of (-) plants w/band	recombination % (A)	recombination % (B)
HpC 39	20/21	10/16	30 (11/37)	62 (10/16)
HpD 9	21/21	16/17	58 (22/38)	6 (1/17)
HpD 23	7/20	9/11	29 (9/31)	18 (2/11)
HpD 39	16/19	12/17	42 (15/36)	70 (12/17)

Recombination percentage A is based on information from both the nod+ and nod- bulks. Recombination percentage B is based on information from only the nod- bulk. Percentages are calculated by adding the number of segregants in each bulk by the total number of plants.

Table (3-8): Polymorphic band distribution from nod139 bulks and parents.

Primer	Band Size (bp)	Presence in *G. soja*	Presence in nod139	Presence in Nod+ bulk	Presence in Nod- bulk
HpA 37	380	Yes	Yes	Yes	No
HpA 40	260	Yes	Yes	Yes	No
HpA 48	300	Yes	Yes	Yes	No
HpA 54	280	Yes	Yes	Yes	No
HpB 12	290	Yes	Yes	Yes	No
HpB 19	240	Yes	Yes	Yes	No
HpB 24	300	Yes	Yes	Yes	No
HpB 30	130	Yes	Yes	Yes	No
HpB 38	400	Yes	Yes	Yes	No
HpB 62	200	Yes	Yes	Yes	No
HpC 39	***600***	***Yes***	***No***	***Yes***	***No***
HpD 7	300	Yes	Yes	Yes	No
HpD 9	***200***	***No***	***Yes***	***No***	***Yes***
HpD 18	500	Yes	Yes	Yes	No
HpD 23	***200***	***No***	***Yes***	***No***	***Yes***
HpD 39	***220***	***Yes***	***No***	***Yes***	***No***
HpD 44	280	Yes	Yes	Yes	No
HpD 49	280	Yes	Yes	Yes	No
HpD 50	400	Yes	Yes	Yes	No
HpD 56	200	No	No	Yes	No
HpD 62	190	Yes	Yes	Yes	No
HpD64	320 &200	Yes, Yes	Yes, Yes	Yes, Yes	No, No
Hp 28	200	Yes	Yes	Yes	No
Hp 38	180	Yes	Yes	No	Yes
Hp 70	100	Yes	Yes	Yes	No
Hp 73	100	Yes	Yes	Yes	No

Bands that proved to be polymorphic in both the bulks and the parents are in bold face and italicized.

Plant #	Nod (+/-)	HpC39 (600bp)	HpD9 (200bp)	HpD23 (200bp)	HpD39 (220bp)	Plant #	Nod (+/-)	HpC39 (600bp)	HpD9 (200bp)	HpD23 (200bp)	HpD39 (220bp)
100	-	+	-	-	+	145	+	+	?	-	+
101	-	+	+	+	+	162	+	+	?	-	+
102	-	+	+	-	+	179	+	+	?	?	?
103	-	-	-	?	-	198	+	+	+	-	+
104	-	+	-	+	+	202	+	+	+	-	+
105	-	+	+	?	+	207	+	+	+	-	?
106	-	+	+	+	+	229	+	+	+	-	-
107	-	-	+	-	+	239	+	+	+	-	+
108	-	+	+	-	-	243	+	+	+	-	+
109	-	-	+	?	+	249	+	+	+	+	-
110	-	-	+	?	+	251	+	+	+	-	+
111	-	+	+	?	+	253	+	+	+	+	?
112	-	-	?	+	+	259	+	+	+	-	?
113	-	+	?	-	-	266	+	+	+	+	?
114	-	-	?	?	+	268	+	+	+	+	?
115	-	-	+	+	-	278	+	+	+	-	?
116	-	+	+	?	-	280	+	-	+	-	?
120	+	+	?	-	?	282	+	+	+	-	?
125	+	+	?	+	+	283	+	+	+	+	?

Table (3-9): Nod139 individual screen results.

(+) indicates nodulation in Nod (+/-) column and presence of band in all other columns. (-) indicates no nodulation in Nod (+/-) column and absence of band in all other columns. (?) denotes missing data due to no amplification or loss of gel, from cracking, at time of table preparation.

Recombination frequencies were also calculated for nod139 which were based only on the nod- bulk data. This was done to eliminate the effect of the heterozygous material in the nod+ bulk. For the majority of the polymorphic bands, this improved the linkage greatly. This suggests that these polymorphisms are actually good markers.

V. Competition experiments

In order to find an explanation for the lack of closely linked markers and to explain the odd nature of the bands found, experiments to evaluate the competition of bands in the DAF reaction were designed. Two separate competition experiments were completed. First, Bermudagrass cv. Tifway (from A. Dudeck, Univ. of Florida, Gainesville) and *G. soja* (PI468.397) templates were mixed in varying ratios by mass, then screened with mini-hairpin primers. The DNA was mixed in the following mass ratios, grass to G. *soja* : 500:1, 100:1, 50:1, 40:1, 30:1, 20:1, 10:1, 5:1, 2:1, 1:1, 1:2, 1:5, 1:10, 1:20, 1:30, 1:40, 1:50, 1:100, and 1:500. The combination of grass and soybean DNA took advantage of the molecular diversity between the two templates by generating a large number of polymorphic bands. This allowed an increase in the information gained from this experiment. Using 9 mini-hairpin primers, 94 polymorphic bands from Bermudagrass, and 62 polymorphic bands from *G. soja* were detected. The raw data of this experiment are in table (3-10). A graphical representation of that data is in figure (3-3). Cumulative data are represented in figure (3-4). A comparison of bands in competition is shown in figure (3-5). This figure demonstrates the variability of competition of bands.

The second set of experiments used G. *soja* vs. nod49 template DNA mixes, which were also screened with mini-hairpin primers. The mass ratios used here were, G. *soja* to nod49: 200:1, 100:1, 20:1, 10:1, 5:1, 1:1, 1:5, 1:10, 1:20, 1:100, and 1:200. The G. *soja* and nod49 mixes were used to more closely resemble the situation we

Table (3-10): Data from competition experiment I (Bermudagrass vs. *G. soja*).

A. Bermudagrass percentage of template

	50	33	20	10	5	3.3	2.5	2	1	0.2
Hp A10	1	2	5	2	2	0	0	0	0	0
Hp A30	2	2	0	2	2	1	0	1	0	1
Hp A50	0	0	0	1	1	0	0	4	0	3
Hp B16	0	1	2	0	2	1	4	0	1	1
Hp B52	0	0	0	3	4	2	0	0	2	0
Hp C19	0	0	0	0	2	0	1	2	2	0
Hp C48	0	0	1	0	3	1	2	1	1	1
Hp D4	0	0	1	2	4	2	0	1	1	0
Hp D60	0	0	1	1	4	2	1	1	0	0
Total # of Bands	**3**	**5**	**11**	**11**	**24**	**9**	**8**	**10**	**7**	**6**

Bermudagrass DNA was diluted with *G. soja* DNA. Values indicate the number of lost polymorphic bands from the grass template at specified dilution.

B. *G. soja* percentage of template

	50	33	20	10	5	3.3	2.5	2	1	0.2
Hp A10	0	1	0	3	2	1	0	2	2	0
Hp A30	0	0	0	2	2	0	1	0	2	0
Hp A50	1	0	3	3	0	0	0	0	0	0
Hp B16	0	4	1	0	1	0	0	0	0	0
Hp B52	0	2	3	2	2	1	0	0	0	0
Hp C19	4	1	0	0	1	1	0	0	0	0
Hp C48	0	1	1	1	0	3	0	0	0	0
Hp D4	0	1	1	1	1	0	0	0	0	0
Hp D60	0	0	0	2	1	1	0	0	0	0
Total # of Bands	**5**	**10**	**9**	**14**	**10**	**7**	**1**	**2**	**4**	**0**

G. soja DNA was diluted with Bermudagrass DNA. Values indicate the number of lost polymorphic bands from *G. soja* template at specified dilution.

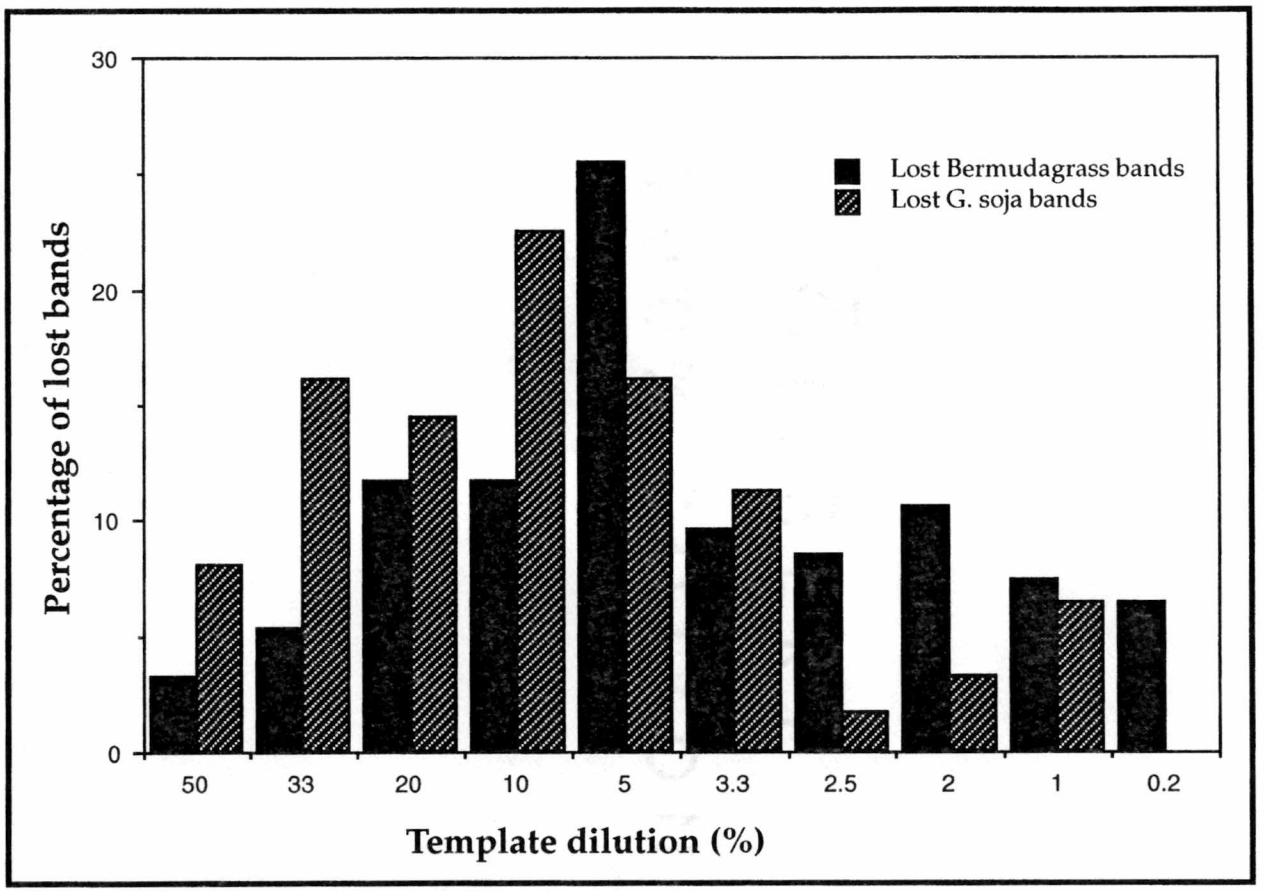


Fig. (3-3) Bermudagrass vs. *G. soja* template competition.

Percentage of lost bands was determined by dividing the number of bands lost at each dilution by the number of total bands lost by that template.

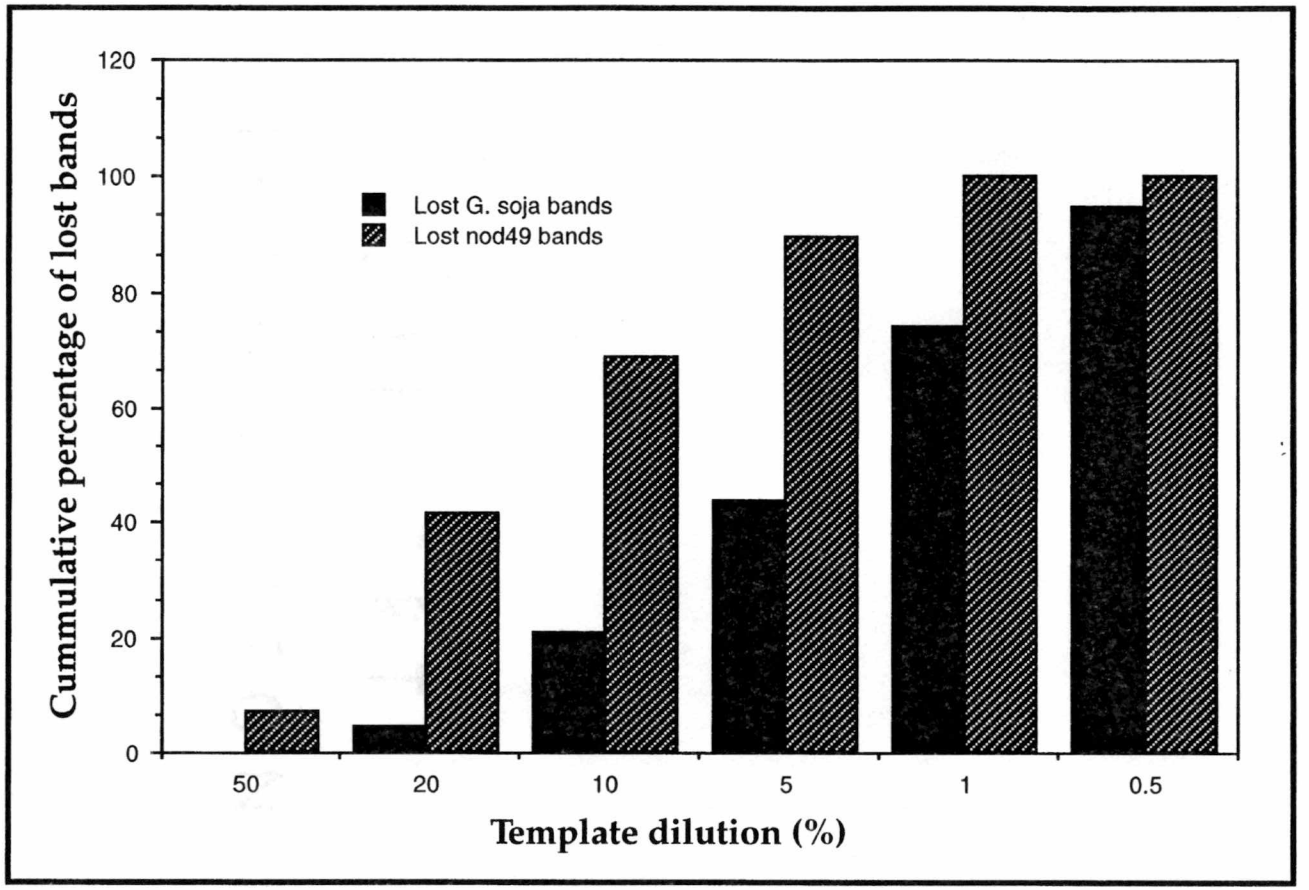


Fig. (3-4) Cumulative percentage of total bands lost in Bermudagrass vs. *G. soja* template competition.

Cummulative percentages are calculated by adding the percentage of bands lost at each dilution to that of all larger dilution percentages for each template.

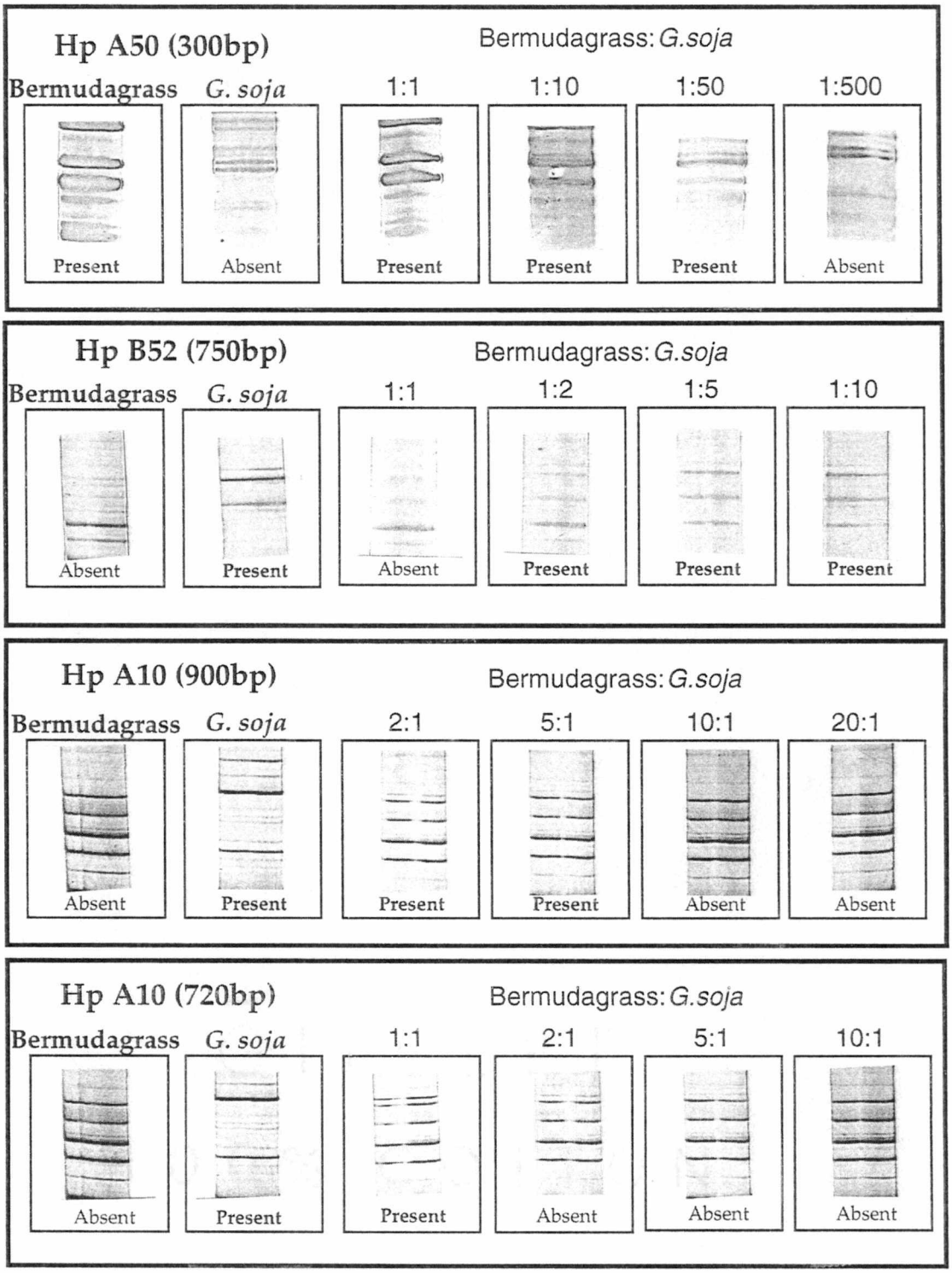


Figure (3-5): Band competition experiment I (Bermudagrass vs. *G. soja*).

These are randomly selected primers that indicate highly variable band competition.

used to screen the nodulation bulks. Using 19 mini-hairpin primers, 43 polymorphic bands from *G. soja*, and 29 polymorphic bands from nod49 were detected. The raw data of this experiment is in table (3-11). A graphical representation of that data is shown in figure (3-6). Cumulative data are represented in figure (3-7). A comparison of bands in competition is shown in figure (3-8). This figure demonstrates the variability of competition of bands.

Table (3-11) Data from competition experiment II (*G. soja* vs. nod49).

	G. soja template: nod49 template										
	200:1	**100:1**	**20:1**	**10:1**	**5:1**	**1:1**	**1:5**	**1:10**	**1:20**	**1:100**	**1:200**
Hp A5			1		1			*1*	*1*		
Hp A28			1		1					*2*	
Hp A37											*1*
Hp A47						1	*1*			*1*	
Hp B19			1	1	1		*1*		*1*	*1*	
Hp B24											
Hp B30			1	1	1			*1*	*1*		*1*
Hp B37		1						*1*			*2*
Hp B42				1				*1*		*1*	*1*
Hp B62		1			1				*2*	*2*	
Hp C39			1		2				*1*		*1*
Hp C51								*1*		*1*	
Hp C10			1	1						*1*	
Hp C19					1	1			*2*		
Hp D43		1		2					*2*		*2*
Hp D60					1			*1*		*1*	
Hp C17				1						*2*	
Hp D1					1			*1*		*1*	*1*
Hp A53				1					*2*		
Total	**0**	**3**	**6**	**8**	**10**	**2**	***2***	***7***	***12***	***13***	***9***

Ratios are given as G. soja : nod49. Italicized and underlined numbers represent bands from *G. soja*, all others from nod49. Values indicate the number of polymorphic bands lost from each template at the specified dilution.

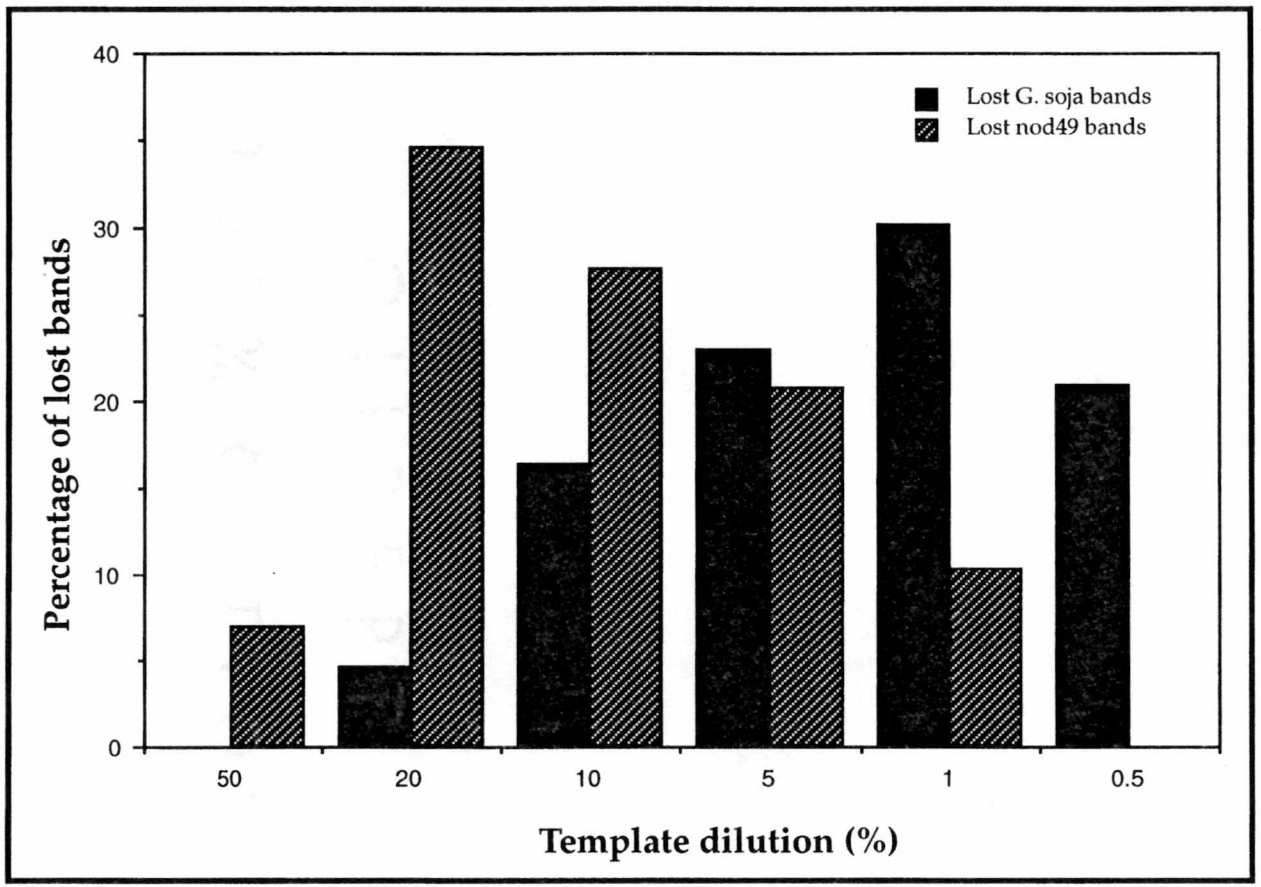


Fig. (3-6) *G. soja* vs. nod49 template competition.

Percentage of bands lost is determined by dividing the number of bands lost at each dilution by the total number of bands lost by that template.

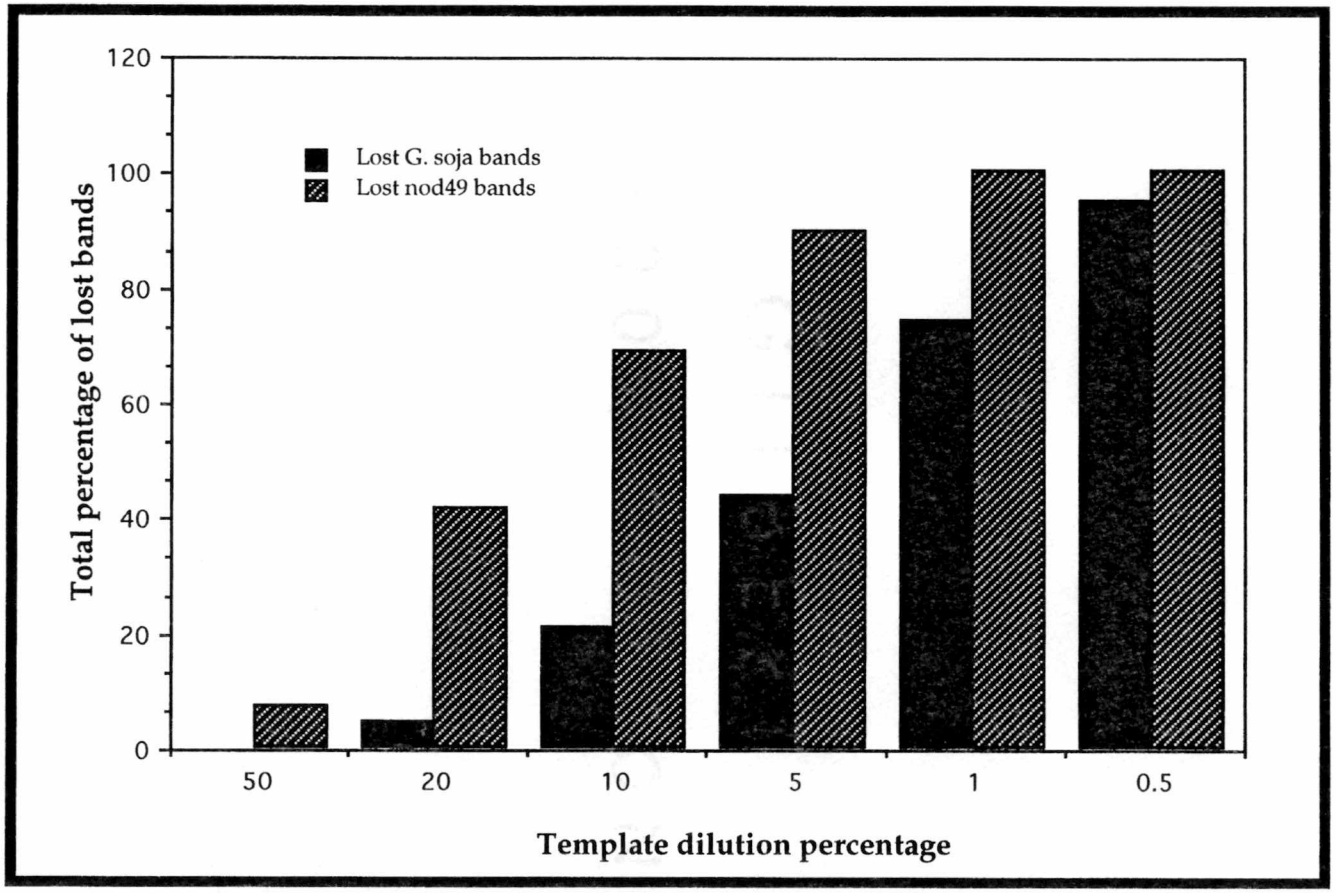


Fig. (3-7) Cumulative percentage of total bands lost in*G. soja* vs. nod49 template competition.

Cummulative percentages are calculated by adding the percentage of bands lost at each dilution to that of all larger dilution percentages for each template.

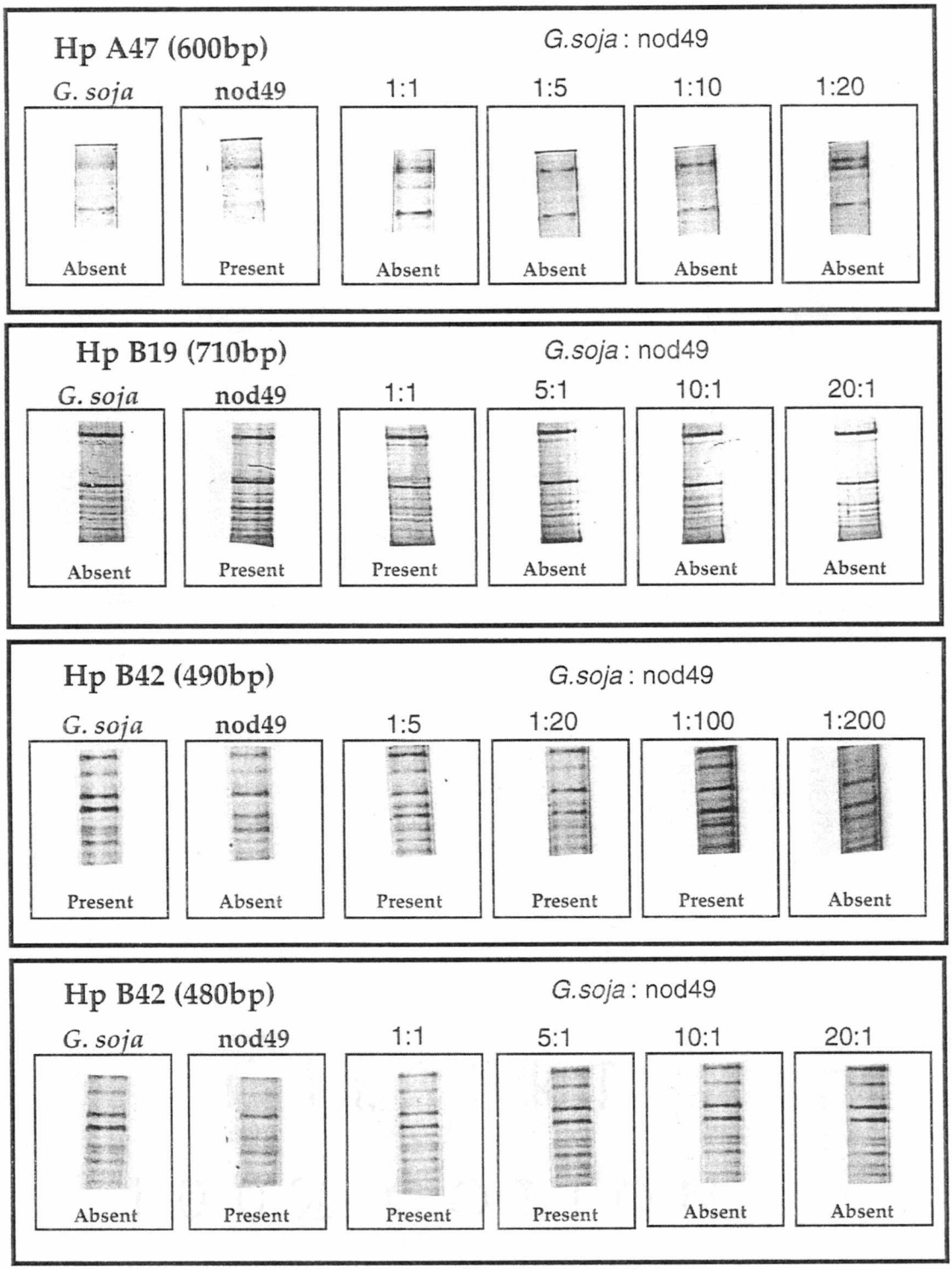


Figure (3-8): Band competition experiment II (*G. soja* vs. nod49).

These are randomly selected primers that indicate highly variable band competition.

Chapter 4
Discussion

I. Primer screening

A screen of 319 mini-hairpin primers was performed on both the nod49 and nod139 bulks with reasonable expectation of finding a closely linked marker. That screen produced 22 nod49 and 27 nod139 potential polymorphic bands. Since nod49 and nod139 are nearly isogenic, a similar number of polymorphisms should be expected from both mutants. Any observable difference may come from the fact that there are two different regions targeted which may have a different polymorphic content. A chi-square analysis of total polymorphic content show that there are no statistical differences, except with the Hp D series. The reliability of these polymorphic bands was then tested in the parental DNA. Those markers that displayed consistency in the parents as well as the bulks were then scored for segregation in the individuals of the F_2 generation. Markers that did not show this consistency were atypical and assumed to be non-inherited. Because of the effort invested in the F_2 screening, this parental pre-screen saved effort by focusing only on the most promising markers. As the data from tables 3-4 and 3-7 indicate, the potential markers that passed the first two screens showed no better than 29% recombination when scored in the F_2 individuals. A typical example of these markers is a 400bp band produced by Hp A47. This band was present in the nodulating bulk of nod49 and absent in the non-nodulating bulk consistently in three separate amplifications. It was, however, present in both the nod49 and the *G. soja* parent. At the time, the behavior of this band was ignored because of its quality and reproducibility. The band was then scored in the F_2 population. Results of the F_2 scoring showed that it displayed 46% recombination. Knowing that 50% recombination indicates independent assortment, it was obvious that this band was not closely linked to the non-nodulating phenotype. Other markers that passed both the bulk

and parental screens gave similar results. Recombination frequencies of these markers ranged from 29% to 46%, which indicate absence of close linkage. When the recombination frequencies were calculated using only data from the nod- bulks, which were known to be homozygous, they did give better values indicating close linkage. But it has not yet been established that data based only on the nod- bulk is valid.

The fact that markers loosely linked were found indicates that the coupling of BSA and DAF works, but inefficiently. These data suggest that the coupling may not increase the ability to isolate genetic markers in soybean as it does with other marker detection methods. In fact, it may be detrimental to the detection of linked markers.

The purpose of the BSA technique is to mask unlinked markers and highlight closely linked ones. In this study, only distant markers were isolated. These findings initially suggested problems in the experimental design. It was apparent that linked markers were either absent or they were being masked in some way. A possible explanation was the existence of competition between bands. Amplification of bands representing linked markers were either masked by the preferential amplification of unrelated bands or, were too competitive and amplified even when rare. Therefore, if linked markers were either very competitive or not competitive at all, they would have a very high or very low probability of being detected in both bulks, respectively. Two competition experiments were performed in order to make a more detailed analysis of the intricacies of DAF that might be affecting the efficiency of BSA.

II. Competition experiments

The first competition experiment consisted of mixing two unrelated templates, bermudagrass and *G. soja*, at different ratios. These ratios varied from 500:1

(grass:soybean) to 1:500, so that the effects of each template on the other would be apparent. This design allowed the study of the competition behavior of a large number of DNA polymorphisms by using material that was genetically distant. This increased the efficiency of the experiment by providing more information per primer. Results of this experiment show that 9 mini-hairpin primers produced 156 polymorphic bands, 94 from bermudagrass and 62 from *G. soja*. The amplification of these bands was very interesting. Some bands were detected when representing only 0.2% of the template concentration, while others were not detected at 50%. The competitive nature of each template was different. Bands from bermudagrass were lost at an average template dilution of 9% (mean = 5%), while bands from *G. soja* were lost at an average of 16% (mean = 10%). This indicated that the bermudagrass template was more competitive than the *G. soja* template. Most of those that were highly competitive were bermudagrass bands while those that showed low competition were mainly *G. soja* bands. These bands demonstrate the wide range of competition possessed by DAF bands.

A second competition experiment was then designed to more closely simulate band competition in the bulks. In this experiment, templates from *G. soja* and *G. max* (nod49) were mixed at ratios varying from 200:1 (*G. soja* :*G. max*) to 1:200. The amount of information per primer was lower in this experiment because the two templates were more closely related. However, more primers were used than in the first experiment in order to increase the data points. A set of 19 mini-hairpin primers were used. These primers produced 72 polymorphic bands, 29 from *G. max* and 43 from *G. soja*. The competitive nature of each template was different. Bands from *G. max* were lost at an average template dilution of 14.2% (mean = 20%), while bands from *G. soja* were lost at an average of 4.4% (mean = 1%). This indicated that the *G. soja* template was more competitive than the *G. max* template.

Other studies have shown differences in band detection. Tinker et al (1994) found a polymorphic RAPD band that was lost at 4.5% of the template. When the same band was used as a Southern analysis probe, it was detectable down to 2.9%. Michelmore et al. (1991) also performed competition experiments. The results of those experiments concluded that RAPD bands lost their detectability at 10%. However, Williams et al (1993) found a large range in the detectability among RAPD bands. Williams et al performed experiments in which template DNA from *Cyanobacterium* and soybean were mixed. Their results showed that all of the bands produced were from the soybean template. Even when the *Cyanobacterium* template was at a 460-fold molar excess, only bands from soybean amplified. Bands from *Cyanobacterium* were completely out competed and did not amplify. Results of these studies indicate that band competition also exists in other marker detection methods.

III. Band competition theories

One conclusion drawn from the competition experiments was that genome size is not a relevant factor of competition. Theoretical formulations would lead one to believe that because one genome is larger, it will have more amplification products and therefore outcompete the smaller genome. This is not true for two reasons. First, the sizes of *G. max* and *G. soja* are similar. Ahmad et al, (1984) reported that the total length of the *G. soja* chromosome complement to be 6-7% smaller than that of *G. max*. Blackhall et al, (1991) used flow cytometry measurements to estimate that the amount of nuclear DNA to be approximately 3% smaller in *G. soja*. Because the *G. soja* template was the most competitive, this data indicates that size is not a deciding factor in competition. Second, DNA profiles of *Escherichia coli, Klebsiella pneumonia, Discula destructiva*, soybean, and human DNA with decamer and octamer primers were similar in the number of amplified

products (Caetano-Anollés, 1994). This indicates that the amount of products produced by a template is not related to genome size, and therefore, that competition is not related to genome size or the number of products generated from each template.

Competition difference in amplification may be the result of a difference in nucleotide sequence variability expressed as genetic diversity between the templates. Differences in genetic diversity between *G. max* and *G. soja* has been established in several studies. Maughan et al, (1995) used a comparison of microsatellite poylmorphisms to establish that the average diversity value for *G. soja* was higher than that of *G. max*. Forty-three more alleles were found in *G. soja* than in *G.max*. Also, Keim et al, (1989) used RFLPs to determine genetic diversity estimates in soybean. Their results indicated that *G. soja* showed higher diversity than *G. max*. Taken together, these observations and the data from the competition experiments suggest that the most competitive template was also the template expressing higher diversity. Since differences in genome size are not followed by changes in the number of amplification products (annealing sites), (Caetano-Anollés, 1994), diversity should affect the quality rather than the number of annealing sites. Theoretically, a template of greater diversity will have a more diverse repertoire of annealing sites. This increases the chances of having sites with high complementarity to the primer. Since a primer is expected to hybridize more efficiently to a highly complementary site than to a site with lower complementarity, primer-template mismatching could be directly related to the competitive ability of an amplification product. Therefore, a genome expressing higher genomic diversity will have a larger repertoire of better annealing sites and will provide a competitive edge during amplification in competition experiments. The number of bands expected from the annealing of an octomer primer to perfectly complementary sites in soybean can be calculated to range from 30 - 68 for primers having fractional GC content of 0.6 - 0.7, (Caetano-Anollés, 1994). DAF

amplifications with octomer primers of soybean DNA, 35 - 39. However, a similar number of bands is also obtained with decamer primers when the expected number of perfectly complementary sites is 0.2. This can only be explained by the existence of primer-template mismatches. Perhaps, genomic diversity is related to the number of highly complementary sites in a genome and this is what causes the observed differences in band competition experiments. DAF has the ability to tolerate mismatching but prefers the amplification highly complementary sites. These conclusions could explain why genomes of high complexity outcompete genomes of low complexity during amplification.

Other explanations of band competition come from the secondary structure of the amplicon. After the first elongation step is completed, a single strand copy of every amplicon is produced. If at that time the single strand copy takes on a non-linear structure, such as a hairpin, the amplification of that product can be inhibited. Amplicons that remain in a linear structure will then have an advantage in the amplification process. By gaining an advantage in the early steps of amplification, amplicons without secondary structure will produce bands that are more competitive.

A final explanation could come from the quality of the DNA used in competition experiments. If one template was more degraded than the other, then the difference in competitiveness of those templates could easily be explained. An agarose gel of the *G. soja* and nod49 DNA show that the level of degradation is close to the same (figure 4-1). The bermudagrass template was unavailable for gel electrophoresis. This gel seems to weaken the degradation argument, but it could still be a factor.

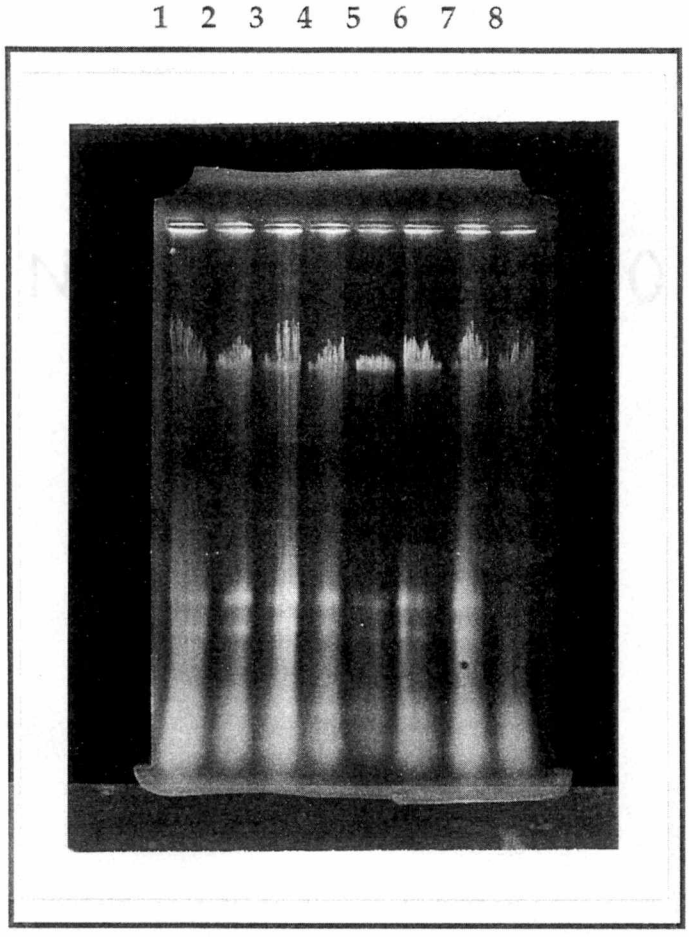


Figure (4-1): Agarose gel electrophoresis of templates.

Lane 1: nod139, Lane 2: nod49, Lane 3: nod49, Lane 4: nod139, Lane 5: *G.soja*, Lane 6: *G. soja*, Lane 7: nod49, Lane 8: nod139. All lanes shows similar levels of degradation.

IV. The lack of closely linked markers

There are several hypotheses that might explain why the primer screen did not result in the isolation of any closely linked marker. They are all plausible and represent good theories to explain this limitation.

First, the region of the genome that has been targeted may simply be sparse in terms of genetic markers. It is known that genetic maps show genetic markers clustered in some areas and not in others. There exists a strong possibility that the regions containing the *nod49* and *nod139* loci lie in areas of conservation where polymorphic content is low. It has been shown that the region around the *nts* locus is conserved (Kolchinsky et al, unpublished). This could show a theme of conservation around the loci that control steps of nodulation. Because these genes are important to the survival of the plant, natural selection would work to keep these regions of the genome secure from mutations. Conservation could be a large factor in explaining these results.

A second hypothesis is that genome duplication could be causing polymorphic bands to appear monomorphic. Because of genome duplication in soybean, the regions around the nod49 and nod139 loci and markers in those regions could be duplicated. If this is true, when the bulks were made, BSA isolated one locus and the surrounding region to one bulk but, the duplicated locus and the region surrounding were present in both bulks (figure 4-2). If this occurs, no markers would be isolated. One marker would be polymorphic between the bulks but, a duplicate copy of that marker would then be monomorphic. The monomorphic marker would mask the detection of the polymorphic marker. The only way to solve this situation would be to create bulks that were homozygous at both loci that control the phenotype. This would require further progeny screening in both nod49 and nod139, and isolation of the duplicate loci in nod49.

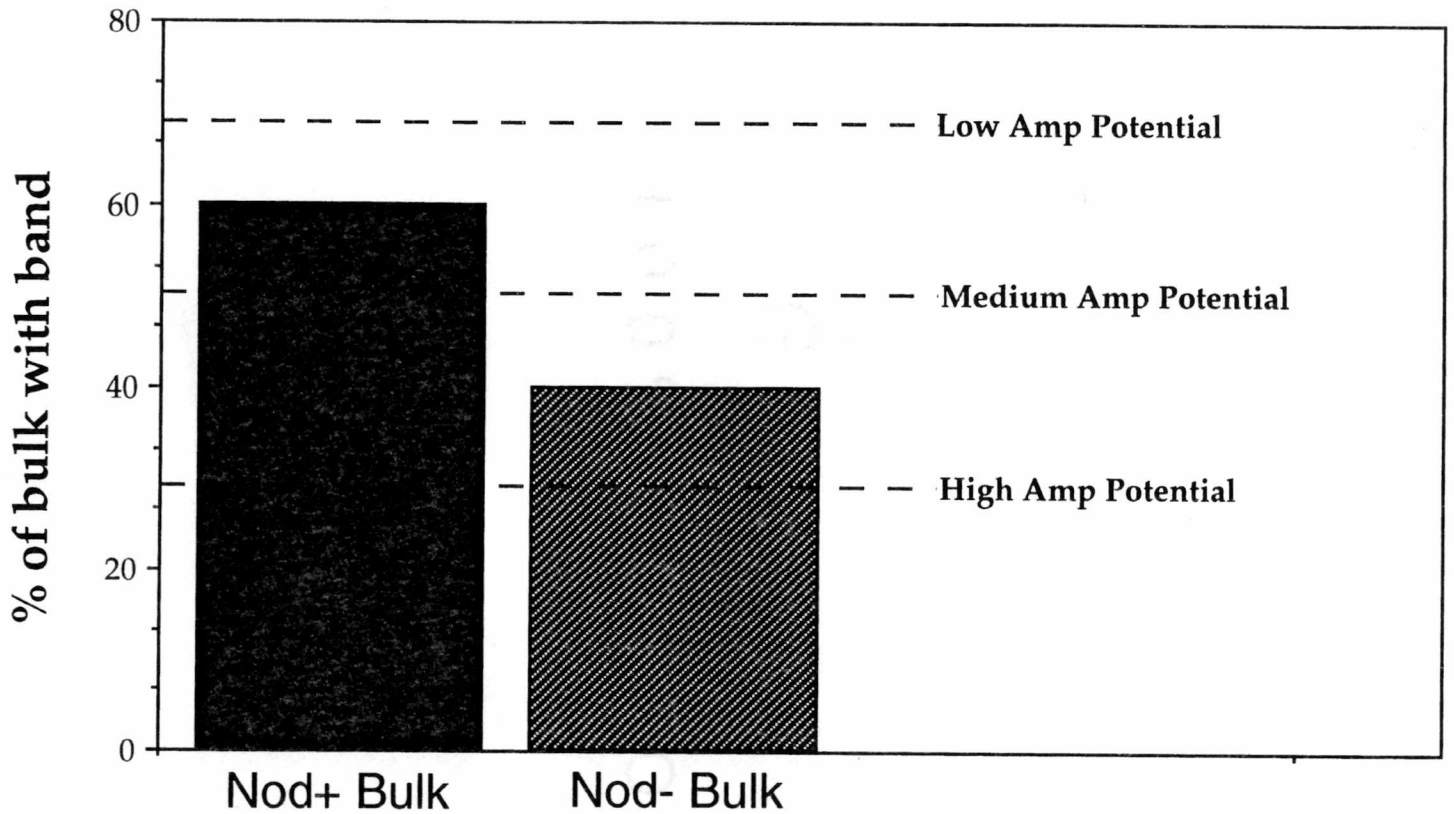


Figure (4-2): Demonstration of amplification potential effects.

The amplification potential of each band is different and it affects the detectability of that band in the bulk. In this case, both high and low amplification potential will result in no polymorphism, but a medium amplification potential will result in the detection of a false polymorphism.

A third explanation, lies in the technical applications of coupling both DAF and BSA together. In this project, the goal was to detect a band in the nod+ bulk and not in the nod- bulk. Most likely, that band would be a dominant marker from the *G. soja* parent. It is however possible to find a co-dominant marker. Detection of a polymorphic marker lies in the ability of the band to be amplified in only one bulk. A term describing the ability of a band to be amplified is "amplification potential" (Tinker et al, 1994). A band with strong amplification potential will be easily amplified while a band with low amplification potential will require better conditions for amplification. Amplification potential is directly related to the competition of potential amplification sites for the primer. A band having a high competitive nature also possesses a high amplification potential, and bands having a less competitive nature will have lower amplification potential.

In a bulking situation, amplification potential and band competition act together with a third variable and control the ability to detect a particular band. The third variable is segregation percentage. The added effect of segregation percentage to amplification potential and band competition can possibly explain all of the irregular marker situations discovered in the screening of the bulks. Consider the following situations: 1) band 1 is present in 60% of bulk A and 40% of bulk B, 2) band 2 in present in 90% of bulk A and 10% of bulk B, and 3) band 3 is present in 10% of bulk A and 1% of bulk B. Band 1 (40% recombination) and band 3 (45.5% recombination) appear to be unlinked, and band 2 (10% recombination) would be linked. In the bulks, band 2 would only be scored as polymorphic if it had amplification potential low enough to inhibit amplification at 10%. Band 3 could be scored as polymorphic if it possessed enough amplification potential to allow amplification at 10% and not at 1%. Band 2 represents most of the markers isolated in the bulk screening. At 40% recombination, this band should theoretically be masked by BSA. However, if the

amplification potential for this band lies between 60% and 40% it will be falsely scored as polymorphic (figure 4-3). The only way a band can be scored as polymorphic is for the threshold of amplification to lie between the segregation percentages of the band in each bulk. That threshold is controlled by band competition and amplification potential. For the case of band 2, if amplification occurs when the band makes up 50% of a bulk, then this band will be detected as polymorphic. It is now apparent that competition experiments should be performed on several of the markers found to determine if amplification potential was the cause of their detection. For example, with the polymorphic band produced by Hp D50 in the nod49 bulks the band is present in 81% of the nod+ bulk and 69% of the nod- bulk. If the band loses amplification between those percentages, this hypothesis will be much stronger.

All of these hypotheses result in a situation where very few markers, if any, can be isolated. Because of genomic conservation, there are few markers in this genomic region already. For those to be detected, their amplification threshold and the segregation percentage have to correlate. This makes it possible for only a small fraction of an already reduced number of markers to be detected.

V. Future directions

Markers for these loci still need to be identified to aid in nodulation research. There are several alternatives that can be applied to fulfill this goal. The first task is to bypass the problems created by genome duplication. This cannot be done using the BSA technique. At this time, it appears impossible to create bulks that are homozygous for both the original and the duplicate locus. In nod139, this cannot be done by further progeny screening. In the F_2, five of the existing nodulating genotypes will not show segregation in any further generation (table 2-2). Also, in nod49 it will not be possible to create this type of bulk since

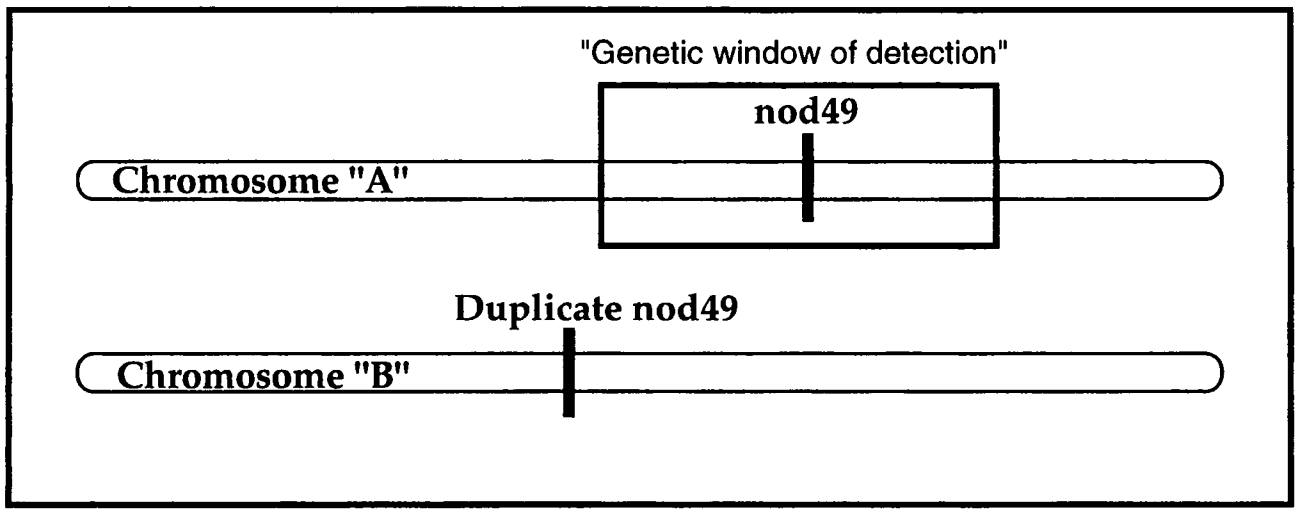


Figure (4-2): The effects of genome duplication on BSA.

The BSA technique isolate a genetic region to one bulk so that polymorphisms may be detected. Everything outside the "window of detection" is present in both bulks equally. When genome duplication has occurred, a duplicate copy of the targeted loci will be present in both bulks. Therefore, detection of the original loci, or any linked markers, is masked by the presence of the duplicate copy.

the duplicated copy of rj_1 has not been identified. Since these homozygous bulks cannot be created, BSA must be left out of the experimental design. The focus should turn to finding polymorphisms between the parents, and then scoring the F_2 population for segregation. Another alternative would be to screen the F_2 population for bands that appear only in the non-nodulating plants (Debbie Landau-Ellis, pers. comm.). This may require the use of a marker detection method that produces a less complex profile, such as RFLP or RAPD. Either of these techniques would circumvent the problems associated with genome duplication.

The second problem to solve is genomic conservation. Since it is thought that these genomic regions are generally conserved and few markers exist, DAF might still be the best detection method to isolate these rare markers. The DAF method has been shown to produce heritable and reproducible molecular markers (Prabhu and Gresshoff, 1994), so it should be possible to isolate DAF markers for these genomic regions by screening more primers. However, using DAF still leaves the problems with band detectability and competition unsolved. The best alternative to DAF would be AFLP (Vos et al, 1995). AFLP shares the DAF characteristic of a large number of bands which indicates a large amount of the genome scanned per primer. The aspect of AFLP that would give it an advantage over DAF is the use of adapter sequences that serve as partially specific primer annealing sites. Since the 5' end of these adapter sequences provide a specific annealing site on every band, annealling would be less arbitrary than DAF and only secondary structure and amplicon length would have an effect on the band competition. The 3' end of the primer would then still have arbitrary characteristics that would keep the level of products high. This might reduce the amount of problems associated with band competition and amplification potential.

VI. Lessons learned

The intricacies of molecular biology can never be taught in a textbook or a classroom. This project appeared to be a stepwise progression leading to several closely linked markers and a short, well defined thesis. However, molecular biology is never a simple process. There are always unseen roadblocks. The purpose of projects such as this are not to make great strides in the genetic engineering race but to decide the best way to win that race. Finding linked markers seemed to be only a matter of working harder and screening more primers. It can now be seen that to win this race, a different method must be employed. Knowing this will now make the isolation of the non-nodulation genes easier.

Literature cited

Ahmad, Q.N., Britten, E.J., Byth, E.D. (1984). The karyotype of *Glycine soja* and its relationship to that of *Glycine max*. Cytologia, 49, 645-658.

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

Barua, U., Chalmers, K., Hackett, C., Thomas, W., Powell, W. and Waugh, R. (1993). Identification of RAPD markers linked to a *Rhynchsporium secalis* resistance locus in barley using near-isogenic lines and bulked segregant analysis. Heredity, 71, 177-184.

Bassam, B. J., Caetano-Anollés, G. and Gresshoff, P. M. (1991). Fast and sensitive silver staining of DNA in polyacrylamide gels. Analytical Biochemistry, 196, 80-83.

Bassam, B. J. and Bentley, S. (1995). Electrophoresis of polyester-backed polyacrylamide gels. Biotechniques, 19, 568-573.

Blackhall, N.W., Hammatt, N., Davey, M.R. (1991). Analysis of variation in the DNA content of *Glycine* species: A flow cytometric study. Soybean Genetics Newletter, 8, 194-200

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

van Brussel, A.A.N., Bakhuizen, R., van Spronsen, P.C., Spaink, H.P., Tak, T., Lugtenberg, B.J.J., and Kijne, J.W. (1992). Induction of preinfection thread structures in the leguminous host plant by mitogenic lipo-oligosaccharides of *Rhizobium*. Science, 257, 70-72.

Caetano-Anollés, G. and Gresshoff, P. M. (1990). Early induction of feedback regulatory responses governing nodulation in soybean (*Glycine Max* (L.) Merr.). Plant Science, 69-81.

Caetano-Anollés, G. and Gresshoff, P. M. (1991a). Excision of nodules induced by *Rhizobium meliloti* exopolysaccharide mutants releases autoregulation in alfalfa. Journal of Plant Physiology, 138, 765-767.

Caetano-Anollés, G., Bassam, B. J. and Gresshoff, P.M. (1991b). DNA amplification fingerprinting using very short arbitrary oligonucleotide primers. Bio/Technology, 9, 553-557.

Caetano-Anollés, G. and Gresshoff, P. M. (1991c). Plant genetic control of nodulation. Annual Review of Microbiology, 45, 345-382.

Caetano-Anollés, G., Joshi, P.A. and Gresshoff, P. M. (1991d). Spontaneous nodules induce feedback suppression of nodulation in alfalfa. Planta, 183, 77-82.

Caetano-Anollés, G., Bassam, B.J., and Gresshoff, P.M. (1992). Primer-template interactions during DNA amplification fingerprinting with single arbitrary oligonucleotides Molecular and General Genetics, 235, 157-165

Caetano-Anollés, G. (1994). MAAP: a versatile and universal tool for genome analysis. Plant Molecular Biology, 25, 1011-1026.

Caetano-Anollés, G. and Gresshoff, P. (1994). DNA Amplification using arbitrary mini-hairpin oligonucleotide primers. Bio/Technology, 12, 619-623.

Caldwell, B. E. (1966). Inheritance of a strain specific ineffective nodulation in soybean. Crop Science, 6, 427-428.

Carlson, R.W., Price, N.P.J., and Stacey, G. (1994). The biosynthesis of rhizobial lipo-oligosaccharide nodulation signal molecules. Molecular Plant-Microbe Interactions, 7, 684-695.

Carroll, B. J. and Gresshoff, P. M. (1983). Nitrate inhibition of nodulation and nitrogen fixation in white clover. Z. Pflanzenphysiology, 110, 77-88.

Carroll, B. J., McNeil, D. and Gresshoff, P. M. (1985). A supernodulation and nitrate tolerant symbiotic (*nts*) soybean mutant. Plant Physiology, 78, 34-40.

Carroll, B. J., McNeil, D. and Gresshoff, P. M. (1986). Mutagenesis of soybean (*Glycine max* (L.) Merr.) and the isolation of non-nodulating mutants. Plant Science, 47, 109-114.

Davis, T. M., Foster, K. W. and Phillips, D. A. (1985). Non-nodulation mutants of chickpea. Crop Science, 26, 719-723.

Dellaporta, S.L., Wood, J., and Hicks, J.B. (1983). A plant DNA mini-preparation. Version II. Plant Molecular Biology Reporter, 1, 19-21.

Devine, T.E., and Breithaupt, B.H. (1980). Significance of incompatibility reactions of *Rhizobium japonicum* strains with soybean host genotypes. Crop Science, 20, 269-271.

Gionannoni, J., Wing, R., Ganal, M. and Tanksley, S. (1991). Isolation of molecular markers from specific chromosomal intervals using DNA pools from existing mapping populations. Nucleic Acids Research, 19, 6553-6558.

Gresshoff, P. M. and Delves, A.C. (1986). Plant genetic approaches to symbiotic nodulation and nitrogen fixation in legumes. In: Blonstein, A.D. and King, P.J. (eds.) Plant Gene Research III. (pp. 159-206). Springer Verlag, Wien.

Gresshoff, P. M. (1993). Molecular genetic analysis of nodulation in soybean. In J. Janick (Ed.), Plant Breeding Reviews, Volume 11. (pp. 275-318). John Wiley and Sons, Inc.

Gresshoff, P. M. (1995). The interface between RFLP techniques, DNA amplification, and plant breeding. In J. Skerrit and R. Apels (eds.), Biotechnology in Agriculture #13: New Diagnostics in Crop Science. (pp.101-125)University Press, Cambridge.

Hellriegel, H. and Wilfarth, H. (1888). Beilageheft, Zeitschrift Ver. Rübenzuckerind Deutschen Reiches.

Hellriegel, H. and Wilfarth, H. (1889). Ber. Deutsch. Bot. Ges., 7, 138-143.

Hirao, I., Nishimura, Y., Tagawa, Y., Wantanabe, K., and Miura, K (1992). Extraordinarily stable mini-hairpins: electrophoretical and thermal properties of the various sequence variants of d(GCGAAGC) and their effect on DNA sequencing. Nucleic Acids Research, 20, 3891-3896.

Hormaza, J., Dollo, L. and Polito, V. (1994). Identification of a RAPD marker linked to sex determination in *Pistacia vera* using bulked segregant analysis. Theoretical and Applied Genetics, 89, 9-13.

Keim, P., Shoemaker, R.C., Palmer, R.G. (1989). Restriction Fragment Length Polymorphism diversity in soybean. Theoretical and Applied Genetics, 77, 786-792.

Kneen, B. E. and LaRue, T. A. (1984). Nodulation resistant mutant of *Pisum sativum* (L.). Journal of Heredity, 75, 238-240.

Kondorosi, A., Kondorosi, E., Banfalvi, Z., Dusha, I., Putnoky, P, and Bachem, C.W.B. (1984). Molecular genetics of symbiotic nitrogen fixation by *Rhizobium meliloti.* Proceedings of the 15th International Congress of Genetics, (pp. 205-216) New Dehli.

Kosslak, R.M., Bookland, R., Barkei, J., Paaren, H., Appelbaum, E. (1987). Induction of *Bradyrhizobium japonicum* common *nod* genes by isoflavones isolated from *Glycine max.* Proceedings of the National Academy of Science USA, 84, 7428-32.

LaFavre, J.S.and Eaglesham, A.R.J. (1984). Increased nodulation of "non-nodulating" (rj_1rj_1) soybeans by high dose inoculation. Plant Soil, 80, 297-300.

Lerouge, P., Roche, P., Faucher, C., Maillet, F., Truchet, G., Promé, J-C., and Dénarié, J. (1990). Symbiotic host-specificity of *Rhizobium meliloti* is determined by a sulphated and acylated glucosamine oligosaccharide signal. Nature, 344: 781-784

Mathews, A., Carroll, B. J. and Gresshoff, P. M. (1987). Characterization of non-nodulation mutants of soybean (*Glycine Max* (L.) Merr.): *Bradyrhizobium* effects and absence of root hair curling. Journal of Plant Physiology, 131, 349-361.

Mathews, A., Carroll, B. J. and Gresshoff, P. M. (1989a). Development of *Bradyrhizobium* infections in supernodulating and non-nodulating mutants of soybean (*Glycine max* [L.] Merr.). Protoplasma, 150, 40-47.

Mathews, A., Carroll, B. J. and Gresshoff, P. M. (1989b). A new non-nodulation gene in soybean. Journal of Heredity, 80, 357-360.

Mathews, A., Landau-Ellis, D., Carroll, B. J., Morgan, J. and Gresshoff, P.(1996). Linkage analysis of two non-nodulation genes in soybean (*Glycine Max* L.) mutants. In Review.

Maughan, P.J., Maroof, M.A.S., Buss, G.R. (1995) Microsatellite and Amplified Sequence Length Polymorphisms in cultivated and wild soybean. Genome, 38, 715-723.

Michelmore, R. W., Paran, I. and Kesseli, R. V. (1991). Identification of markers linked to disease resistance 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.

Mohan, M., Nair, S., Bentur, J., Rao, U. and Bennett, J. (1994). RFLP and RAPD mapping of the rice *Gm2* gene that confers resistance to biotype 1 of gall midge (*Orseolia oryzae*). Theoretical and Applied Genetics, 87, 782-788.

Mulongoy, K., and Akobundu, I.O. (1990). Agronomic and economic benefits of nitrogen contributed by legumes in live mulch and alley cropping systems. In Gresshoff, P., Stacey, G. (eds.), Nitrogen Fixation: Achievements and objectives. Proceedings of the 8th International Congress on Nitrogen Fixation, Knoxville, TN USA. p. 625. Chapman Hall, New York London (1990).

Peters, K., Frost, J., and Long, S.R. (1986). A plant flavone, luteolin, induces expression of *Rhizobium meliloti* nodulation genes. Science, 233, 977-80.

Prabhu, R.R., and Gresshoff, P.M. (1994). Inheritance of polymorphic markers generated by DNA amplification fingerprinting and their use as genetic markers in soybean. Plant Molecular Biology, 26, 105-116.

Pracht, J. E., Nickell, C. D. and Harper, J. E. (1993). Genes controlling nodulation in soybean: rj_5 and rj_6. Crop Science., 33, 711-713.

Quispel, A. (1988). Nitrogen Fixation: Hundred Years After. In H. Bothe, F. J. de Bruijn and W. E. Newton (Eds.), Proceedings of the 7th International Congress on Nitrogen Fixation. (pp. 3-12). Gustav Fischer, New York..

Redmond, J.W., Batley, M., Djordjevic, M.A., Innes R.W., Kuempel, P.L., and Rolfe, B.G. (1986). Flavones induce expression of nodulation genes in *Rhizobium*. Nature, 323, 632-35.

Relic, B., Fellay, R., Lewin, A., Perret, X., Price, N.P.J., Rochepeau, P., and Broughton, W.J. (1993). *Nod* genes and Nod factors of *Rhizobium* species NGR234. New Horizons in Nitrogen Fixation , (pp. 183-189), Palacios, R., Mora, J., and Newton, W.E. (eds). Kluwer Academic Publishers, Dordrecht.

Relic, B., Perret, X., Estrada-Garcia, M.T., Kopcinska, J., Golinowski, W., Krishnan, H.B., Pueppke, S.G., and Broughton, W.J. (1994). Nod factors of *Rhizobium* are a key to the legume door. Molecular Microbiology, 13: 171-178.

Rolfe, B. G. and Gresshoff, P. M. (1988). Genetic analysis of legume nodule initiation. Annual review of Plant Molecular Biology, 39, 297-319.

Stacey, G., Minami, E., Kouchi, H., Cohn, J.R., Carlson, R.W. (1995). Signal exchange during soybean nodulation by *Bradyrhizobium japonicum*. In Tikhonovich, I.A., Provorov, N.A., Romanov, V.I., Newton, W.E. (eds.), Nitrogen Fixation: Fundamentals and Applications. (pp 305-310). Kluwer Academic publishers, , Dordrecht.

Sutherland, T.D., Bassam, B.J., Schuller, L.J., and Gresshoff, P.M. (1990). Early nodulation signals of the wild type and symbiotic mutants of soybean (*Glycine max*). Molecular Plant-Microbe Interactions, 3, 122-128.

Timmerman, G., Frew, T., Weeden, N., Miller, A. and Goulden, D. (1994). Linkage analysis of *er-1*, a recessive *Pisum sativum* gene for resistance to powdery mildew fungus *(Erysiphe pisi* D.C.). Theoretical and Applied Genetics, 88, 1050-1055.

Tinker, N. A., Mather, D. E. and Fortin, M. G. (1994). Pooled DNA for linkage analysis: practical and statistical considerations. Genome, 37, 999-1004.

Truchet, G., Barker, D.G., Camut, S., de Billy, F., Vasse, J., and Huguet, T. (1989). Alfalfa nodulation in the absence of *Rhizobium*. Molecular and General Genetics, 219, 65-68.

Truchet, G., Roche, P., Lerouge, P., Vasse, J., Camut, S., de Billy, F., Promé, J-C., and Dénarié, J. (1991). Sulphated lipo-oligosaccharide signals of *Rhizobium meliloti* elicit root nodule organogenisis in alfalfa. Nature, 351: 670-673.

Vest, G. (1970). *Rj3*, a gene conditioning ineffective nodulation in soybean. Crop Science, 10, 34-35.

Vest, G. and Caldwell, B. E. (1972). *Rj4* , a gene conditioning ineffective nodulation in soybean. Crop Science, 12, 692-693.

Vos, P., Hogers, R., Bleeker, M., Reijans, M., van de Lee, T., Hornes, M., Fritjers, A., Pot, J., Kuiper, M., and Zabeau, M. (1995). AFLP: a new concept for DNA fingerprinting. Nucleic Acids Research, 8, 785-794.

Welsh, J. and McClelland, M. (1990). Fingerprinting genomes using PCR with arbitrary primers. Nucleic Acids Research, 18, 7213-7218.

Wicking, C. and Williamson, B. (1991). From linked marker to gene. Trends in Genetics, 7, 288-290.

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

Williams, J., Reiter, R., Young, R. and Scolnik, P. (1993). Genetic mapping of mutations using phenotypic pools and mapped RAPD markers. Nucleic Acids Research, 21, 2697-2702.

Williams, L. F. and Lynch, D. L. (1954). Inheritance of a non-nodulating character in the soybean. Agronomy Journal, 46, 28-29.

Xie, Z-P., Staehelin, C., Vierheilig, H., Wiemken, A., Vögeli-lange, R., Broughton, W.J., and Boller, T. (1995). Nodulation factors stimulate the mycorrhizal colonization of nodulating and non-nodulating soybeans. Plant Physiology, 108, 1519-1529.

Vita

Shane Abbitt was born in Bloomington, Indiana on July 26, 1971. He graduated from Edgewood High School in Ellettsville, Indiana in May, 1989. He entered Indiana University, Bloomington, Indiana in August, 1989 then transferred to Simpson College, Indianola, Iowa in August, 1990 where he received the Bachelor of Arts in Biology in May 1993. He then entered the M.S. in Biotechnology program at the University of Tennessee, Knoxville, Tennessee in August, 1993 where he received his degree in August, 1996. On June 3, 1995, he was married to Tara Lynn Allen in Crossville, Tennessee.

He is presently pursuing opportunities with DNA fingerprinting in the fields of research and forensics.