To the Graduate Council:

I am submitting herewith a dissertation written by Rekha Rathnakar Prabhu entitled "Molecular Mapping in Soybean Using DNA Amplification Fingerprinting." I have examined the final copy of this dissertation for form and content and recommend that it be accepted in partial fulfillment of the requirements for the degree of Doctor of Philosophy, with a major in Life Sciences.

Dr. Peter M. Gresshoff, Major Professor

We have read this dissertation
and recommend its acceptance:

Accepted for the Council:

Associate Vice Chancellor and
Dean of The Graduate School

MOLECULAR MAPPING IN SOYBEAN USING DNA AMPLIFICATION FINGERPRINTING

A Dissertation
Presented for the
Doctor of Philosphy
Degree
The University of Tennessee, Knoxville

Rekha Rathnakar Prabhu
August 1995

DEDICATION

This thesis is dedicated to my parents

Mr. Rathnakar Srinivas Prabhu

and

Mrs. Sarita Rathnakar Prabhu

who gave invaluable support, guidance and opportunity for my education. I am grateful for this.

ACKNOWLEDGMENTS

First and foremost, I would like to thank my advisor, Dr. P. M. Gresshoff, for giving me the opportunity to obtain a doctoral degree. He has been very positive and encouraging to me. I have enjoyed the stimulating academic environment in his laboratory which has enhanced scientific experimentation and learning. I am grateful to my committee members for many helpful suggestions which made a positive contribution to the evolution of my dissertation.

There have been several people in the laboratory who have been helpful and I thank them for their friendship and support. In particular, I would like to mention Debbie Landau-Ellis for helping me with the plant material and with whom I have developed a special friendship that I hope will endure. I would also like to acknowledge Janice Crockett for her prompt and efficient help and practical advice.

I would like to thank Amy MacKenzie for preliminary work on the Pioneer material and the group at Pioneer Hi-Bred, Drs. D. Webb, S. Smith, S. Luk and H. Jessen, for providing the plant materials, RFLP pedigree data, and cluster analysis. I would like to especially thank Dr. Gordon Lark for supplying recombinant inbred DNA and analysis of mapping data. Dr. Alexander Kolchinsky is thanked for providing the impetus for the collaboration with Dr. Lark and for his always helpful advice on scientific matters.

This work was supported by a grant from the Tennessee Soybean Promotion Board, the Human Frontiers Science Programme, the United Soybean Board and the Ivan Racheff Endowment.

ABSTRACT

This study dealt with the molecular genetic analysis of soybean (*Glycine max* (L.) Merr.) using predominantly single primer amplification markers. The long term aim was to provide additional information needed for the cloning of genes involved in the process of nodulation in legumes. Initial emphasis was given to the mutation nts382 which confers supernodulation and the lack of autoregulation to soybean. The *nts* mutation as well as the onset of autoregulation are not associated with major polypeptide changes in the leaves of soybean. DNA amplification fingerprinting (DAF) using high primer to template ratio and single, very short arbitrary primers, was used to generate DAF markers in soybean (*Glycine max* (L.) Merr.). The amplification reaction was carried out with soybean genomic DNA and 8 base long oligonucleotide primers. The DAF reaction was optimized for template and enzyme concentration. Silver stained 5% (w/v) polyacrylamide gels containing 7 M urea detected between 11 and 28 DAF products with primers of varying GC content (ranging from 50-100% GC). Depending on the intensity, DAF markers were classified into three classes. The intensity of silver-stained DNA fragments was linear from 150 pg to 5 ng/band and did not show any abnormal mobility for size or amount of DNA in each band. DAF profiles were reproducible for different DNA extractions and gels. Forty markers were detected by 26 primers when comparing *Glycine soja* and *Glycine max*. The inheritance of DAF markers was studied using a cross between the ancestral *Glycine soja* PI468.397 and *Glycine max* (L.) Merr. line nts382, F1, and F2 progeny. Most DAF markers were inherited as dominant Mendelian markers in F1 and F2 populations. However, abnormal inheritance occurred with about 25% of the polymorphisms. One marker was inherited as a maternal marker, presumably originating from organelle DNA while another showed apparent paternal inheritance. To confirm the nuclear origin and utility of dominant Mendelian markers, three DAF polymorphisms were mapped using a F11 mapping population of recombinant inbred lines from soybean cultivars 'Minsoy' x 'Noir1'. The utility of DAF markers was also demonstrated for pedigree assessment and cultivar identification. DAF data produced a dendrogram similar to that produced by RFLP and pedigree data. Cloning and hybridization of DAF products showed that DAF polymorphic bands may contain sequences of multiple origin. The study showed that DAF-generated polymorphic markers occur frequently and reliably, that they are inherited as Mendelian dominant loci and that they can be used in genome mapping.

TABLE OF CONTENTS

LIST OF TABLES

TABLE PAGE

PART IV: INHERITANCE OF POLYMORPHIC FRAGMENTS GENERATED BY DNA AMPLIFICATION FINGERPRINTING AND THEIR USE AS GENETIC MARKERS IN SOYBEAN

PART V: USE OF DNA AMPLIFICATION FINGERPRINTING FOR GENOTYPE DISTINCTION AND PEDIGREE ASSESSMENT OF SOYBEAN CULTIVARS

PART VI: GENOMIC BASIS OF DNA AMPLIFICATION MARKERS

LIST OF FIGURES

FIGURE PAGE

PART I: GENERAL INTRODUCTION AND LITERATURE REVIEW

PART II: POLYPEPTIDE PATTERNS IN LEAVES OF *GLYCINE MAX* (L.) MERR. CV. BRAGG AND ITS SUPERNODULATING MUTANT NTS382 DURING EARLY NODULATION

PART III: DETECTION OF DNA POLYMORPHISMS BY DNA AMPLIFICATION FINGERPRINTING

PART IV: INHERITANCE OF POLYMORPHIC FRAGMENTS GENERATED BY DNA AMPLIFICATION FINGERPRINTING AND THEIR USE AS GENETIC MARKERS IN SOYBEAN

PART V: USE OF DNA AMPLIFICATION FINGERPRINTING FOR GENOTYPE DISTINCTION AND PEDIGREE ASSESSMENT OF SOYBEAN CULTIVARS

PART VI: GENOMIC BASIS OF DNA AMPLIFICATION MARKERS

LIST OF ABBREVIATIONS

AFLP	Amplified Fragment Length Polymorphism
AP-PCR	Arbitrarily Primed-Polymerase Chain Reaction
APT	Arbitrary Primer Technology/ies
AU	Absorbance Unit/s
BAC	Bacterial Artificial Chromosome
CAPS	Cleaved Amplified Polymorphic Sequence
CGE	Capillary Gel Electrophoresis
DAF	DNA Amplification Fingerprinting
DAI	Days After Inoculation
DGGE	Denaturing Gradient Gel Electrophoresis
GMS	Genomic Mismatch Scanning
HPLC	High Pressure Liquid Chromatography
LOD	Likelihood Difference
MLG	Molecular Linkage Group
NOR	Nucleolus Organizer Region
NPA	N-(1-naphthyl) phthalamic acid
PCR	Polymerase Chain Reaction
QTL	Quantitative Trait Loci
RAPD	Random Amplified Polymorphic DNA
RDA	Representation Difference Analysis
RFLP	Restriction Fragment Length Polymorphism
RT-PCR	Reverse Transcriptase-Polymerase Chain Reaction
RUBISCO	Ribulose 1,5-Bisphosphate Carboxylase/Oxygenase
SCAR	Sequence Characterized Amplified Region
SDI	Shoot Derived Inhibitor
SRFA	Selective Restriction Fragment Amplification
SSCP	Single Strand Conformational Polymorphism
SSR	Simple Sequence Repeat
STR	Short Tandem Repeat
TAS	Telomere Associated Sequence
tecMAAP	Template Endonuclease Cleaved Multiple Arbitrary Amplicon Profiling
VNTR	Variable Number Tandem Repeat
YAC	Yeast Artificial Chromosome

LIST OF ABBREVIATIONS

AFLP	Amplified Fragment Length Polymorphism
AP-PCR	Arbitrarily Primed-Polymerase Chain Reaction
APT	Arbitrary Primer Technology/ies
AU	Absorbance Unit/s
BAC	Bacterial Artificial Chromosome
CAPS	Cleaved Amplified Polymorphic Sequence
CGE	Capillary Gel Electrophoresis
DAF	DNA Amplification Fingerprinting
DAI	Days After Inoculation
DGGE	Denaturing Gradient Gel Electrophoresis
GMS	Genomic Mismatch Scanning
HPLC	High Pressure Liquid Chromatography
LOD	Likelihood Difference
MLG	Molecular Linkage Group
NOR	Nucleolus Organizer Region
NPA	N-(1-naphthyl) phthalamic acid
PCR	Polymerase Chain Reaction
QTL	Quantitative Trait Loci
RAPD	Random Amplified Polymorphic DNA
RDA	Representation Difference Analysis
RFLP	Restriction Fragment Length Polymorphism
RT-PCR	Reverse Transcriptase-Polymerase Chain Reaction
RUBISCO	Ribulose 1,5-Bisphosphate Carboxylase/Oxygenase
SCAR	Sequence Characterized Amplified Region
SDI	Shoot Derived Inhibitor
SRFA	Selective Restriction Fragment Amplification
SSCP	Single Strand Conformational Polymorphism
SSR	Simple Sequence Repeat
STR	Short Tandem Repeat
TAS	Telomere Associated Sequence
tecMAAP	Template Endonuclease Cleaved Multiple Arbitrary Amplicon Profiling
VNTR	Variable Number Tandem Repeat
YAC	Yeast Artificial Chromosome

PART I

GENERAL INTRODUCTION AND LITERATURE REVIEW

I: BIOLOGICAL NITROGEN FIXATION

Nitrogen is a major limiting factor in the growth of a plant. Elemental nitrogen is a very important constituent of amino acids, pyrimidines and purines as well as many other essential compounds where it occurs in the oxidation state, -3. Various free-living prokaryotic organisms such as cyanobacteria, soil organisms e.g., *Klebsiella*, *Azotobacter*, *Frankia*, anaerobic photo-autotropic bacteria e.g., *Clostridium*, and *Desulfovibrio* have the ability to utilize N_2 gas as a source of nitrogen [Postgate 1982]. Although eukaryotic plants do not have the ability to fix atmospheric N_2 *per se*, they can acquire this ability through symbiosis with specialized prokaryotic organisms [Parker 1957].

Symbiotic nitrogen fixation was first conclusively demonstrated by Hellriegel and Wilfarth in 1888 [Hellriegel and Wilfarth, 1888]. In an outstanding piece of work that involved a comparative study of the relationship between added nitrogen and the yield of nitrogen in a variety of plants, they observed that the leguminous plants reacted differently compared to all other plant groups. They concluded that legumes could use atmospheric nitrogen and that this capability was associated with root nodules induced and occupied by bacteria. In a symbiotic relationship, in return for reduced N_2, the plants supply the bacterium with carbon (photosynthate). The process of reducing N_2 to NH_3 by bacteria is termed "Biological Nitrogen Fixation". The enzyme responsible for fixing atmospheric nitrogen called nitrogenase, is found only in prokaryotes where it functions under strict anaerobic conditions [Eady et al., 1979].

Legumes such as soybean, pea, clover, bean, and alfalfa form symbiotic associations with soil bacteria of the genus *Rhizobium*, *Bradyrhizobium*, and *Azorhizobium* [Bauer 1981]. The plant family Leguminosae is the third largest family in the angiosperms and is found in a variety of habitats ranging from the arctics to the tropics. This successful diversity is owed in part by the legumes to their ability to become autotropic for fixed nitrogen. Non-legume symbioses occur between a tree called *Parasponia* and *Bradyrhizobium* [Trinick 1979], actinomycetes of the genus *Frankia* and Alder, and between cyanobacteria and lichens, mosses, ferns and marine worms [Torrey 1987]. *Azotobacter* and *Azospirillum* are free-living N_2-fixers that participate in "associative N_2 fixation" with roots of C4 grasses [Dobereiner et al., 1976].

The knowledge gained from studying biological nitrogen fixation for more than a hundred years has resulted in agricultural practices that are both economically attractive and

ecologically sound. In an agriculturally sustainable system, nitrogen fixing legumes and non-legumes contribute to processes that regenerate resources. Rotation or inter-cropping of legumes with other crops, planting nitrogen fixing trees to enrich soil and provide firewood, alley cropping, use of cover crops, and green manure are a few examples of such processes [Mulungoy and Akobundu 1990, Bohlool 1990]. Real life examples of the benefits of these practices are numerous. In field trials with soybean at five diverse sites, results show that the yield of soybean increased by about 28% with N fertilizer compared with inoculation with *Rhizobium*. However, the economic return was almost three times higher with inoculation than with fertilizer [Bohlool 1990]. The *Azolla-Anabena* symbiosis has been used as a nitrogen source in rice paddies as well as a source of food for fish [Lumpkin 1987]. *Sesbania rostrata* is a fast-growing, high N_2-fixing legume with potential as a green manure (nitrogen source) for rice [Ladha et al., 1990]. Free-living organisms such as *Clostridium, Azotobacter* and *Azospirillum* have been extremely beneficial in improving nitrogen content in grasslands and farms e.g., *C. pasteurianum* was estimated to add around 1 kg N per hectare per year which is significant when compared to that fixed by red clover which is around 300 kg N per hectare per year [Postgate 1982].

II: NODULATION

Overview of nodulation

Soybean, peanut, mungbean, cowpea are nodulated by *Bradyrhizobium* spp. and form determinate nodules. Pea, clover, and *Medicago* spp. including alfalfa on the other hand, are inoculated by *Rhizobium* spp. and form indeterminate nodules [Nap and Bisseling 1990]. The distinction between determinate and indeterminate nodule type is based on the presence of a true meristematic zone in the indeterminate but not the determinate nodule. The two types of rhizobia exhibit differences with respect to a range of characteristics e.g., host range, rate of growth, exopolysaccharide composition, and intrinsic antibiotic resistance. Furthermore, there are distinct genetic differences such as the separation of nitrogenase *nif HDK* genes in *Bradyrhizobium* and the presence of symbiotic plasmids in *Rhizobium* but not *bradyrhizobium*. [Martinez et al., 1990].

A variety of molecular, genetic, biochemical, and cell biological studies have revealed three sequential stages leading to the development of a nodule; (1) nodule initiation and root hair curling, (2) infection and (3) nodule development culminating in nitrogen-fixing symbiosis. A schematic diagram of the steps involved in soybean nodule initiation and development is

shown in Fig. 1.

Plant-bacteria interactions begin with chemical signaling in the rhizosphere. The molecules originating from the plant that act as chemoattractants of rhizobia have been identified as flavones, isoflavones and isoflavonoids. Rhizobia secrete substances called Nod factors that induce root hairs to branch, deform and curl resulting in entrapment of bacteria in a pocket of root cell walls. A variety of signals exchanged between legumes and bacteria result in nodule initiation. In soybean, this step is required for penetration and infection and occurs within 12 hours after contact between root and bacteria. Infection threads grow through the root hair cell and into the root cortex where the bacteria are released into the host cytoplasm. At this stage the bacteria are enclosed in a host-derived peribacteroid membrane and later differentiate into bacteroids with the onset of nitrogen fixation. A more extensive description of nodulation can be found in reviews by Appelbaum [1990] and Hirsch [1992].

Role of the plant in nodulation

Alfalfa and white clover, are capable of producing nodules in the absence of *Rhizobium*. These nodules are referred to as spontaneous (if produced spontaneously) [Caetano-Anollés et al., 1991c, Truchet et al., 1989, Blauenfeldt et al., 1994, F. Ligero pers. comm.], or referred to as pseudonodules (if produced by the action of chemicals such as triiodobenzoic acid or NPA N-(1-naphthyl) phthalamic acid [Hirsch et al., 1989]). This suggests that the *Rhizobium* modulates fundamental capacities already present in the plant. A number of plant genes are induced upon contact of host root with rhizobia, e.g., nodulins, conventionally defined as proteins expressed specifically in the nodule. Nodulins are involved in nodule organogenesis, maintenance, and function. Recent studies however, suggest the expression and role of nodulins in other cellular processes. For example, *Ps*ENOD12 (pea) and *Ms*ENOD12 (*Medicago sativa*) have been detected in stems and flowers [Scheres et al., 1990, Franssen et al., 1990, Bauer et al., 1994], leghemoglobin transcripts can be detected as early as two days after inoculation suggesting constitutive expression [F. de Bruijn pers. comm.]. *ENOD40* in pea and soybean is found in uninoculated roots and stems, and has been suggested to function as an RNA molecule due to the absence of an open reading frame and the unusual stability of its RNA, characteristic of non-coding RNAs [Matvienko et al., 1994, Kouchi and Hata 1993, Crespi et al., 1994]. It should be noted that none of the nodulins have yet been shown to be required for nodulation or nitrogen fixation. In other words, no symbiotically altered mutation has been

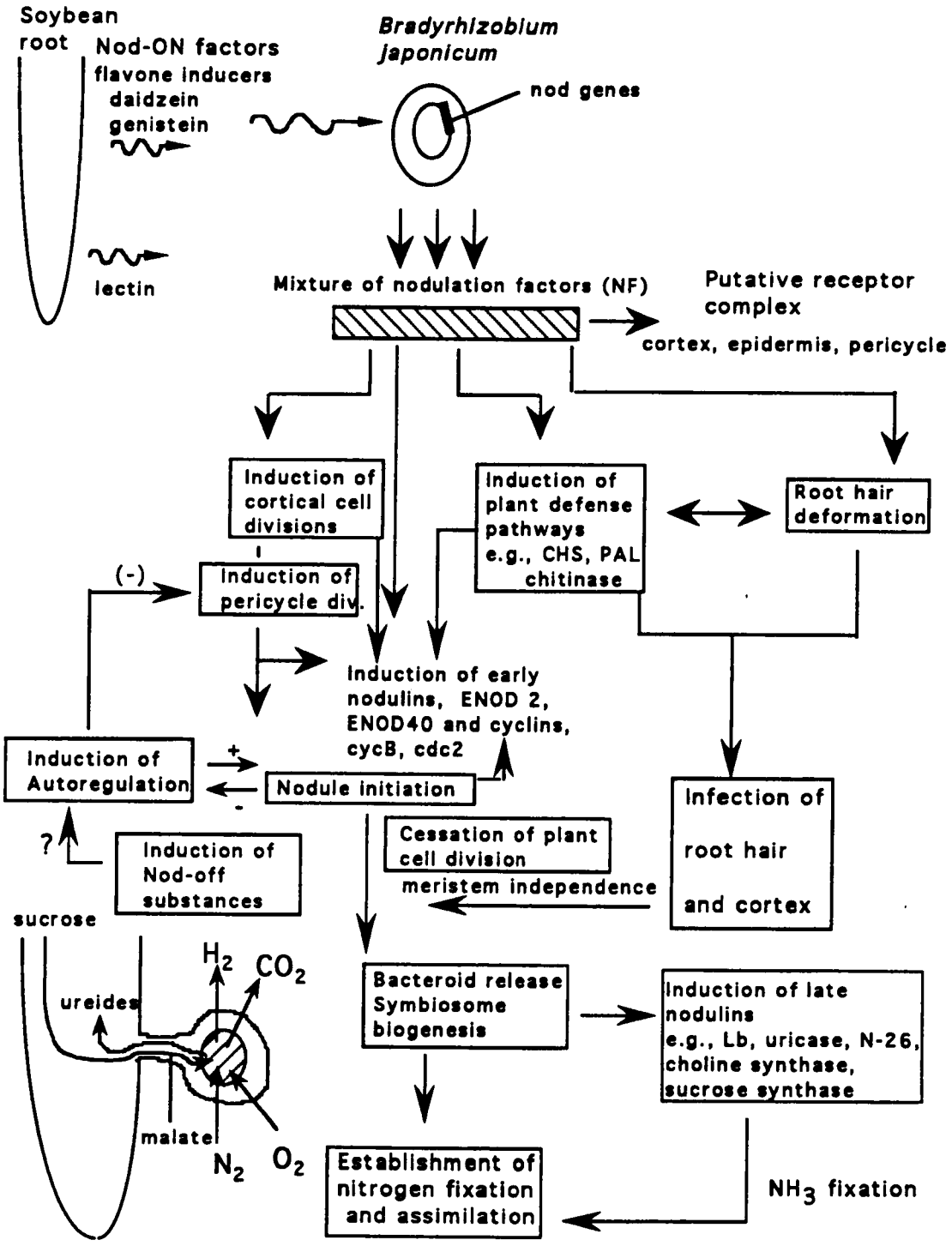


Figure 1: Schematic diagram showing nodule initiation and development

correlated to a nodulin, and no nodulin has been mutated *in vitro* and replaced in plants. Verma and associates used antisense mutagenesis in an attempt to verify the essential nature of some nodulins, but data were inconclusive possibly due to the effect of transformation [Lee et al., 1993].

Based on the time point of expression during nodule development, nodulin genes have been divided into early and late genes. The early genes are already expressed before the start of nitrogen fixation, while the expression of late genes is first detectable around the onset of nitrogen fixation. In addition to the chemoattractants mentioned above, a significant number of symbiotic functions are under the direct control of the plant. For example, the site of infection, regulation of nodule number, nitrate-sensitivity to nodulation, the amount of oxygen and carbohydrates that the bacteria receives, and the termination of symbiosis through the senescence of the plant and regulation of nodulation.

Identifying plant genes in nodulation

Conclusive proof of the role of the plant in symbiosis can only be achieved through genetic and molecular biology techniques. Several strategies have been used to identify plant loci involved in symbiosis. Differential screening, (N-35, uricase [Suzuki and Verma 1991]) subtractive hybridization (ENOD40, [Kouchi and Hata 1993]) and EMS mutagenesis (see below). Currently, methods of differential display [Liang and Pardee 1993] and insertional mutagenesis with T-DNA [Walden et al., 1991] or transposons [Chuck et al., 1993] are being used with some success. Techniques like subtractive hybridization and differential display (the latter especially) are much more sensitive and can detect low abundance mRNAs. For example clones homologous to chitinases have been identified in *Sesbania rostrata* by differential display [F. de Bruijn pers. comm.]. Genes that are common to nodules and roots are not likely to be detected by subtractive and differential display techniques, a limitation that is overcome by tagging techniques. Gene tagging methods cause insertional mutagenesis to produce phenotypes, therefore new genes can be identified. Although transposon tagging methods have the disadvantage of producing mutant phenotypes, these methods allow direct isolation of the genes by cloning DNA flanking the transposon. Transposon tagged genes can be isolated by inverse PCR [James et al., 1995] and T-DNA tagged genes can be cloned using plasmid rescue [Koncz et al., 1989] and linker mediated PCR [Mueller et al., 1992].

Chimeric genes and transgenic plants are used to study regulation and function of

nodulation genes. Two plants that currently enjoy the status of "model legume" are *Lotus japonicus* [Handberg and Stouggard 1992] and *Medicago truncatula* [Barker et al., 1990]. Among several advantages of growth and propagation, symbiosis-deficient mutants can be generated in these legumes, the mutations can be genetically mapped, the plants can be transformed with *Agrobacterium tumefaciens* and *Agrobacterium rhizogenes* and fully regenerated plants can be recovered. In addition, these organisms have a relatively small genome size (per haploid genome, *L. japonicus*, 0.5 x 10^9 bp; *M. truncatula*, 1.8 x 10^9 bp) and are nodulated. *A. rhizogenes* induces hairy roots in plants and therefore T-DNA can be used to introduce DNA and to yield transgenic nodules when infected with *Rhizobium* [Hansen et al., 1989].

Antisense methods can be used to determine gene function. Recent examples include infection with *A. tumefaciens* bearing antisense constructs of *M. truncatula* ENOD40. Transgenic plants showed arrested callus growth while overexpressing *Mt*ENOD40 embryos developed teratomas. These data suggest a role for ENOD40 in growth and differentiation [Crespi et al., 1994]. The antisense construct of N-35 from *Vigna aconitifolia* shows a reduction in ureide biosynthesis affected the development of peroxisomes and resulted in nitrogen deficiency symptoms in plants [Lee et al., 1993]. Gene fusion systems like the soybean *lbc3* promoter fused to CAT can be expressed in a heterologous system like alfalfa or *Lotus corniculata* to study the role of *cis* or *trans* acting elements in leghemoglobin gene expression [F. de Bruijn et al., 1989]

III: SYSTEMIC REGULATION OF NODULATION IN LEGUMES

Supernodulating and nonnodulating mutants in soybean

Through chemical mutagenesis (EMS) several supernodulating and nonnodulating mutants of soybean were isolated by Gresshoff and colleagues in 1985 [Carroll et al., 1985 a,b; Carroll et al., 1986]. Naturally occurring nonnodulating variants have been observed in several legume species like clover, soybean, pea, peanut, alfalfa, chickpea [Caetano-Anollés et al., 1991b]. In soybean, Williams and Lynch [1954] identified rj_1, a monogenic recessive, naturally occurring gene for nodulation. Since then, several other loci, Rj_2, Rj_4, rj_5 and rj_6, have been shown to be involved in soybean nodulation [Caldwell 1966, Vest 1970, Vest and Caldwell 1972, Mathews et al., 1989, Pracht et al., 1993]. The number of available mutants exhibiting non- or ineffective nodulation ranges from two in peanut to over twenty in pea. However, mutants that nodulate profusely in the presence of nitrate

have been identified in pea, soybean, and bean [Caetano-Anollés et al., 1991b] only through mutagenesis of these legumes.

EMS mutagenesis produced three nonnodulating mutants defining two loci that control early nodulation in Bragg. Nod139 (rj_6) fails to show sub-epidermal cell division and root hair curling and has a single gene mutation which is recessive, nuclear and governs non-nodulation. Nod49, a non-nodulator, is allelic to rj_1 but complements nod139 to give wild-type phenotype [Mathews et al., 1987]. Nod49 and nod772 lack root hair curling but induce cell divisions and both nod139 and nod49 epistatically suppress the Nts phenotype [Mathews et al., 1989].

Twelve soybean mutants which exhibited increased ability to nodulate in the presence of nitrate (Nts phenotype) were produced by EMS mutagenesis. All twelve mutants belong to the same complementation group despite having differing phenotypes in their degree of tolerance to nitrate [Delves et al., 1988]. The supernodulating mutants, nts382 and nts1007 develop about 30 times more nodules than wild-type Bragg, and nts1116 has 2-5 times the nodule number/plant. Other phenotypes like altered nodule morphology, lowered nitrogenase activity, reduced root growth, altered ratio of root to shoot growth and nitrate -tolerant nodulation were observed in nts382 and was associated with the original selection of the supernodulating phenotype [Carroll et al., 1985a].

Nitrate inhibition of nodulation

Nodulation is tightly controlled in the presence of nitrate and the number of infections exceeds the number of nodules formed. Inhibition by nitrate has been observed at several stages of symbiosis namely, attachment of rhizobia to the root [Dazzo and Brill 1978], root hair formation, root hair curling [Munns 1968], infection thread formation [Malik et al., 1987], and the level of immunologically detectable lectin on the surface of the root [Sherwood et al., 1984]. Nitrate also delays the appearance of nodules [Gibson and Harper 1985], inhibits the number of nodules that are formed, inhibits specific nitrogenase activity and nodule development, and induces early senescence [Carroll et al., 1985b].

Autoregulation

Autoregulation of nodulation is characterized as control of nodule size, number, and distribution by the host plant. The existence of the phenomenon of autoregulation and its shoot-controlled, systemic nature came from a variety of experiments [Bhuvaneswari 1981,

Pierce and Bauer 1983]. Nodule excision experiments in soybean and alfalfa by Caetano-Anollés et al., [1991b] showed that nodule primordia are arrested at an early stage by older nodules. In soybean, new nodules were developed in the region of initial nodulation between the root tip and the emerging root hairs at the time of inoculation. In alfalfa however, it was observed that no nodules were found in the initial window of nodule initiation, suggesting that alfalfa controls nodulation earlier than soybean and at the onset of cell division [Caetano-Anollés and Gresshoff 1991a,b]. Both alfalfa and soybean control nodulation through the control of cell division, the point of control precedes nitrogen fixation, and is independent of the nitrogen status of the plant. The work of Bauer and co-workers [Calvert et al., 1984] showed that prior inoculation of a region on the root prevented nodulation in the younger root tip region even with a second inoculation. This indicated that non-emergent nodules controlled the fate of pre-nodules. This type of autoregulation was later shown to exist in siratro [Ridge and Rolfe 1988], subterranean clover [Sargent et al., 1987], *Parasponia* [Bender and Rolfe 1988] and alfalfa [Caetano-Anollés and Bauer 1988, Caetano-Anollés and Gresshoff 1991a].

The fact that autoregulation defective mutants are also nitrate tolerant suggests that these two processes are coordinately controlled. Supernodulating mutants nodulate more quickly than the wild-type Bragg and form more nodules above the root tip unlike Bragg. According to a model suggested by Carroll and Mathews [1990], the autoregulation inhibitor and nitrate slow the rate of development of infections in wild type. These two components act in a synergistic manner in that the presence of nitrate decreases the rate of formation of infection threads, thereby reducing the actual number of infection centers. In supernodulating mutants the autoregulation inhibitor is diminished, and since nitrate cannot exert its effect without the inhibitor according to the hypothesis, this results in nitrate-tolerant nodulation. This also allows infection centers to be developmentally committed resulting in increased number of nodules. These explanations are consistent with observations that in the presence of nitrate, infection thread formation does not solely limit nodulation and that if roots were exposed to nitrate after a 18 hour delay, the inhibition of nodulation was considerably relieved. Nodule growth is not inhibited in supernodulating mutants, suggesting that developmentally committed nodules are dependent on the autoregulation inhibitor alone.

Mutants nts382, and nts1007 are blocked at the autoregulation stage which in wild-type plants tends to arrest the symbiotic cell division at a stage in nodule development (stages are defined in Mathews et al., [1989] and Gerahty et al., [1990]. In wild-type Bragg the

number of actual infections and pseudo-infections per centimeter of root is decreased significantly after stage III (about 48 hrs after inoculation) of nodule development [Mathews et al., 1989]. The transition of both types of nodule initiation foci, pseudoinfections and actual infections to advanced nodule development is restricted. In contrast, in mutant nts382 this transition (beyond stage III and IV) is less restricted compared to wild-type resulting in the supernodulating phenotype. All these observations suggest that (1) the regulation of nodulation is internal to the nodule and (2) nitrate inhibition of early nodulation in soybean involves the autoregulation step within about 48 hours after inoculation [Gresshoff et al., 1989].

Experiments using the split root system made important contributions towards identifying the "transmission of an inhibitor" and not "depletion of a critical molecule" as the cause for nodulation inhibition. In a study by Hinson [1975] one half of the split-root system was exposed to nitrate while the other half had none. Even though both sides were inoculated at the same time, nodule numbers on the nitrate treated side were significantly lowered while the untreated side had control level of nodule numbers. Prior inoculation on one side of a soybean split root system [Kosslak and Bohlool 1984] suppressed nodulation on the other side as early as 24 hours after prior inoculation [Olsson 1988]. A 7 day delayed inoculation of Bragg showed complete suppression on the delayed root portion whereas nts382 did not show this suppression [Olsson et al., 1988]. Using the split root system of the supernodulating mutant, in the presence of low and non-inhibitory levels of nitrate, time-separated inoculation allowed both systems to develop well [Olsson et al., 1989]. These studies (a) also showed the systemic nature of shoot controlled-autoregulation and (b) demonstrated for the first time that autoregulation was involved in the suppression of nodulation and not preferential allocation of shoot photosynthate as hypothesized earlier [Singleton and Kessel 1987].

Evidence for the direct involvement of the shoot in autoregulation came from another set of experiments involving reciprocal grafts between the epicotyl and hypocotyl of wild-type (Bragg), Nts and Nod^- plants. Those grafts with nts382 epicotyl and Bragg hypocotyls supernodulated and showed nitrate tolerance whereas the reverse grafts did not. All supernodulating mutants acted through the shoot whereas the nonnodulating phenotype was root-controlled and the nts382 shoot did not suppress this phenotype [Delves et al., 1986, 1987a,b]. To resolve the production of an inhibitor in wild-type Bragg or the presence of an activator in supernodulating mutant, 'challenge' experiments were designed. Shoots from nts or wild-type plants whose roots had been inoculated (challenged) first,

were grafted to uninoculated roots. The roots of the grafts were then inoculated. Bragg shoots lowered nodule numbers after grafting with increased time of challenge, whereas nts382 did not show any increased nodulation. This suggested that the signal for autoregulation was an inhibitor in wild-type Bragg while nts382 lacked this inhibitor.

Evidence for early induction of a feedback regulatory response came from approach graft experiments by Caetano-Anollés and Gresshoff [1990]. Two plants whose genotypes were identical or different were grafted just below the cotyledons to give a fused plant with two shoot and two root systems. A 72 hour delay in inoculation on one root of a Bragg iso-graft resulted in suppression of nodulation whereas no corresponding suppression was seen in nts382 grafts. When nod139 which fails to induce cortical cell divisions was grafted on Bragg or nts382, no suppression of nodulation occurred. Nod49 however, which induces some cortical cell divisions that do not develop beyond the very early stages of nodule formation, did show systematic suppression. These results showed that cortical cell division foci in early stage of nodule development elicit feedback suppressive responses in wild-type but these responses are defective in supernodulating mutant nts382.

Serial sectioning of Bragg and nts382 showed that Bragg had an abundance of early stage nodules while nts382 showed nodules at later stages IV and V [as defined by Calvert et al., [1984] of development [Mathews et al., 1989]. The work of Gerahty et al., [1991] confirmed that Bragg roots developed stage IV pre-nodules within 48-72 hours. Gresshoff and Caetano-Anollés [1991b] therefore suggest that the onset of autoregulation takes place within 72 hours and that a "wave" of the autoregulation signal affects the root 5-7 days after inoculation.

Pea mutant nod_3 [Jacobsens 1984], and soybean mutant nts382, are both characterized by increased lateral root formation, indicating a pleiotropic effect of these mutations and a possible hormonal imbalance in these plants. Supernodulation is controlled by the root in the nod_3 mutant [Postma et al., 1988], but it is controlled by the shoot in the pea mutant isolated by Messager [1985]. This suggests that defects in both root and shoot can cause supernodulation at least in the case of pea.

Model for autoregulation of nodulation

General model for autoregulation

First proposed by Gresshoff and Delves [1986], a general model for the regulation of

nodulation is described (see part II, Fig. 1) in which the first cortical cell divisions induced by *Bradyrhizobium* produce a signal "Q" [Caetano-Anollés and Gresshoff 1990]. Upon translocatation of Q to the shoot, a second signal, the "shoot derived inhibitor" (SDI) is produced. SDI arrests further cortical cell divisions in the root by suppressing nodulation in the ontogenetically younger root tissues. This feedback regulation of nodule number is operative in wild type, but is defective in the supernodulating nts382.

The involvement of phytohormones

Experiments by Olsson [1988], showed that phytohormones are involved in the systemic communication between shoot and root for control of nodulation. Their function is complex and other physiological parameters like the nitrogen status of the plant play a part in phytohormone-mediated regulation of nodulation. In order to accommodate these findings and based on recent developments in the field, Gresshoff proposed an "auxin burst control" hypothesis [Gresshoff 1993]. The main premise of the hypothesis is that the plant responds to a change in auxin transport by increasing auxin derived from the shoot to the roots. The change in auxin transport is caused presumably by the nod-factors (from the bacteria) that elicit flavone changes in the epidermis, root hairs and cortex resulting in inhibition of radial auxin transport from the phloem parenchyma. Smit et al., [1993] identified a pea stele factor which induced cortical cell divisions in the root cortex similar to the position of nodules in inoculated roots. This factor has now been purified and shown to be uridine [J. Kijne pers. comm.]. This "burst" of auxin to the root results in blocking or slowing of further cell division or reinitiation of cell divisions in younger tissues. As a result, the soybean nodule grows by cell expansion only, induced by auxin. Supernodulation in soybean is caused by a decreased ability of the mutant to generate this burst of auxin. The ability of the supernodulating mutant to nodulate faster can be explained in terms of having a lower threshold of auxin levels and suggests a better "readiness" to nodulate. If nitrate is presumed to increase sensitivity to auxin, it could explain the initial inhibition of nodulation effects in nitrate treated plants and also explain the nitrate tolerant phenotype of nts mutants.

IV: NODULATION-A MODEL SYSTEM

Nodulation research in a legume like soybean or pea is an excellent system for studying applied and fundamental aspects of plant biology. The legume-(*Brady*)*rhizobium* interaction involving complex interactions for nodule initiation and development provides a system for studying a myriad of cellular processes: signal transduction, study of regulation

of cortical cell divisions providing insight into cell cycle genes and their control, study of plant gene expression with host genes involved in autoregulation, study of nodulation as a developmentally regulated phenomenon, and study of plant immune response in the context of symbiosis. Understanding the function of agronomically important genes like *nts* is invaluable for introgression into other varieties or for transformation into species that do not fix nitrogen.

Nodulation mutants have been invaluable in delineating the sequence of events in autoregulation thereby allowing better understanding of the function and phenotype of the mutated gene. However, they have not produced any information on the structure of the gene. The problem of knowing a gene only through its phenotype is found in many different areas of plant biology, e.g., resistance genes and morphological mutations. A strategy for gene isolation that does not require prior knowledge of the gene product is a map-based or positional cloning strategy that has been successful for cloning genes from humans, higher plants, yeast, and fruitflies.

Map-based cloning

Maps required for cloning of genes

A detailed genetic analysis in any organism is made possible by a primary genetic linkage map consisting of easily scored polymorphic loci spread throughout the genome. Classically, linkage maps have been constructed in organisms such as bacteria, yeast and fruitflies in which many visible mutations were available as genetic markers. The availability of an abundant supply of polymorphic markers like isozymes and RFLPs (restriction fragment length polymorphisms) (see below) has made possible the construction of linkage maps in many other organisms like humans, yeast, viruses, bacteria and *Drosophila* and several plant species such as soybean, maize, tomato, wheat, lettuce, barley and *Arabidopsis* [O'Brien 1993]. A milestone in molecular mapping was the publication of the first molecular map of the human genome by Donis-Keller et al., [1987].

The availability and use of molecular markers has had a tremendous impact in the area of human genetics. Efforts to improve quality of human life has focused research on localization of human disease genes. More than 400 rare human diseases exhibiting simple Mendelian behavior have been mapped; of these, forty have been cloned e.g., Down's syndrome, Huntington's disease, and muscular dystrophy [Cooper and Schmidtke 1992, Koenig et al., 1987]. Some "complex genes" i.e., those genes that do not exhibit classic

Mendelian behavior have also been cloned or mapped, e.g., cystic fibrosis [Riordan et al., 1989, Rommens et al., 1989] and breast cancer susceptibility [Wooster et al., 1994]. Molecular genetic analysis of model organisms such as mouse, *Drosophila*, *C. elegans*, and *S. cerevisae* have provided extensive knowledge that has been used in the mammalian genome and its manipulation. It was proposed in 1980 that systematic genetic mapping of the human genome be undertaken in an unprecedented international collaboration known as the Human Genome Project (HGP). Currently, the mean interval between the various physical and genetic maps is 0.7 cM [Murray et al., 1994].

Genetic and physical maps yield qualitatively different information. Genetic markers link the entire genome and provide an overall view of the entire chromosome with a relatively low level of resolution whereas physical maps provide very high resolution of certain regions without necessarily linking very distant regions. A physical map provides information regarding the actual distance in base pairs between two linked markers. A physical map is extremely advantageous for two important reasons. The map serves as a vehicle for progression from a genetically defined locus to actually cloning the gene. If the locus has been genetically mapped, the physical map can be used to isolate overlapping clones which encompass the locus of interest. Secondly, a physical map provides a starting point for studying "whole genomic organization" e.g., the physical linkage of cloned genes and the distribution and organization of repetitive elements. The map provides a framework for correlating physical and genetic distances and for integrating information obtained at the nucleotide level as in the case of sequence information.

There are two types of physical maps, macro-restriction maps and ordered clone maps. Macro-restricted maps are generated by using infrequently cutting restriction enzymes to produce a set of large DNA fragments. The order of the fragments is deduced by analyzing the pattern and size of the restricted fragments. This type of a map provides information at the level of the intact chromosome thereby providing long range continuity in the organization of the DNA fragments. An ordered clone map consists of an overlapping collection of clones which can be obtained in a variety of vectors like YAC (Yeast Artificial Chromosome) [Burke et al., 1987], cosmid (a lambda derived plasmid vector for cloning large inserts), phage [Ausubel et al., 1994] or BAC (Bacterial Artificial Chromosome) [Shizuya et al., 1992]. The major advantage of this type of map is the high resolution they provide along with the ready availability of clones.

YACs are produced by a partial digestion of genomic DNA. The resulting restriction (e.g.,

*Eco*RI) fragments are size selected on a pulse-field gel before ligating to the vector, pYAC4. The insert is ligated between the arms of the vector which contains yeast ARS elements, centromeres, telomeres and selectable markers for propagation as an artificial chromosome when transformed into yeast. YACs present problems for chromosome walking in terms of chimeras, deletions, and the presence of repetitive sequences at their termini. To overcome some of these problems BAC libraries are being constructed for a variety of species like soybean and rice. BACs contain smaller insert sizes, average insert size in the BAC library is 125 kb as for example in the rice library [Wang et al., 1995]. However, larger human genome fragments, >300 kb fragments have been shown to be clonable [Shizuya et al., 1992]. YAC libraries have the advantage of containing large fragments (order of Mb) compared to cosmid or phage P1, and have been constructed or are under construction for a number of crops e.g., soybean [P. M. Gresshoff, in progress] tomato [Martin et al., 1992], rice [Ashikawa et al., 1992] and *Arabidopsis* [Ward and Jen 1992].

Relationships among maps

If phenotypic or isozyme markers have already been assigned to specific chromosomes, relating genetic and cytogenetic maps is relatively straightforward. *In situ* hybridization is a convenient method to locate clones (RFLP, YACs etc.) on specific chromosomes utilizing aneuploids such as monosomics and trisomics, and substitution lines. Use of substitution lines similar in concept to aneuploid lines and is commonly used in cereals. Lines with known chromosomes or chromosome arms are substituted with homeologous segments from alien species such as in wheat and rye addition lines. Each substitution line has a different extra chromosome and the clone can be easily localized on the chromosome. Trisomic lines (2n=2x+1) have a different extra chromosome and are useful for associating linkage groups with specific chromosomes.

Genetic maps are not as saturated as they could be mainly for two reasons, the existence of multiple maps for a given species and proprietary information about markers. Competing efforts have led to the production of multiple maps each based on a different cross. This means that information of the location of a gene on one map may not be transferable to another map. This is because markers that are polymorphic in one mapping population may not be in another. For example, RFLP marker pUTG-132a, (see below) tightly linked to the *nts* locus was polymorphic between *G. soja* and *G. max* but not between *G. max* cultivars 'Minsoy' and 'Noir 1' (D. Landau-Ellis pers. comm.). This resulted in mapping

nts to the soybean map constructed from a cross involving *G. soja* and *G. max* (see below) but could not be mapped to the map constructed using *G. max* cultivars Minsoy and Noir 1 (see below).

Synteny

Molecular markers have found use in a recent important observation that many distantly related plant species have co-linear maps for portions of their genome. First demonstrated for members of the Solanaceae family, tomato and potato [Bonierbale et al., 1988], extensive co-linearity has been observed between maize and sorghum [Whitkus et al., 1992], wheat and rice [Kurata et al., 1994], as well as rye and barley [Devos et al., 1992]. Even more distantly related species might be related as has been shown in the case of the duplicated sucrose synthase loci found on the same linkage group in both maize and wheat. These findings have important applications in prediction of gene order and linkage groups in a species based on the order of mapped gene loci in another species provided that the colinearity that exists at the macro-level (tens of megabases) extends to the micro-level (tens of kilobases) e.g., the structural gene loci encoding the monomeric isoenzymes of NADH dehydrogenase have been located on chromosome 4 of *Triticum* species and rye [Figueiras et al., 1991]. Comparison of maps from species containing homeologous chromosomes allows the discussion of structural differences in the genome, map lengths, and genomic distribution of chromosomal regions where markers show distorted segregations. Genome conservation has been shown for three legume genera, *Vigna*, *Phaseolus* and *Glycine* [Boutin et al., 1994]. Two other important applications are the additional resource of molecular markers from another genera or species e.g., markers from mungbean and *Phaseolus vulgaris* have been assigned to the *nts* region in soybean (see below), a 92 base repeat has been shown to occur in *G. max* and *Vigna unguiculata* (cowpea) [Kolchinsky and Gresshoff 1995], tomato markers have been mapped in potato [Gebhardt et al., 1991] and a sorghum map constructed with maize markers [Ragab et al., 1994]. Lastly, information obtained from smaller genomes like rice and sorghum can be used to map genes in larger genomes [Ahn et al., 1993].

Cloning genes of interest

It is a major challenge to clone genes whose function is not known. Map-based cloning of such genes involves a series of steps. The locus of the trait of interest is first genetically mapped. Once closely linked markers flanking the gene of interest have been identified, the

next task is to bridge the gap between the markers. These tightly linked markers (close to 0% recombination) could still be located millions of base pairs away. The genetic markers are then used to identify YAC (or other cloned genomic libraries) in which they are contained. Overlapping YACs are identified using end-clones. The walk is continued until a YAC contig is found that bridges the gap between the two markers. This kind of approach for cloning genes known only by phenotype therefore utilizes a variety of powerful new techniques of RFLP [Botstein et al., 1980], PCR [Mullis et al., 1986], APT* (Arbitrary Primer Technology) [Caetano-Anollés et al., 1991b; Williams et al., 1990; Welsh and McClelland 1990, Bassam and Bentley 1994], PFGE [Schwartz and Cantor 1984], and YAC cloning methods.

The success of a map-based cloning approach will depend on several factors. A high density map i.e., low kbp/cM as in *Arabidopsis* (290 kbp/cM) [Arumuganathan and Earle 1991, Leutwiler et al., 1984] tomato (1 marker/cM) [Tanksley 1994] and rice (265 kbp/cM) [Ronald 1993]; small genome size, e.g., *Arabidopsis* (70 Mbp) and rice (450 Mbp); low frequency of interspersed repeats; the availability of YAC or other large insert DNA libraries. The tomato genome has approximately 80% of its genome as single copy DNA [Zamir and Tanksley 1988], major repetitive elements have been characterized [Ganal et al., 1988] with a detailed knowledge of the long range structure of telomeres [Ganal et al., 1991]. All these factors have contributed to the success of map-based cloning in tomato e.g., the *Pto* resistance gene to *Pseudomonas syringae* [Martin et al., 1993] and the *Fen* gene controlling sensitivity to fenthion [Martin et al., 1994]. Other plant genes that have been cloned using positional cloning include the blight disease resistance locus *Xa21* in rice [Ronald et al., 1992], *AB13* gene controlling decreased responsiveness to abscissic acid in *Arabidopsis* [Giraudat et al., 1992], the *ag* (agamous) [Yanofsky et al., 1990] gene in *Arabidopsis* and the *ap3* (apetala 3) gene in *Arabidopsis* [Jack et al., 1992]. Recent examples of agronomically important cloned genes include the map-based cloning of the *RPS2* gene in *Arabidopsis* conferring resistance to the bacterial pathogen *Pseudomonas syringae* [Bent et al., 1994].

Recent developments indicate that "chromosome walking" is lengthy and tedious especially with larger genomes. Problems associated with repetitive sequences and chimerism as in YAC clones make walking difficult since repetitive DNA makes identification of end clones difficult. Recent strategies are moving away from this approach towards a more "zero in"

* The term APT was first used in Bassam BJ, Bentley, S: DNA fingerprinting using Arbitrary Primer Technology (APT): A tool or a torment. Aust Biotech 4: 232-236 (1994).

or "chromosome landing" approach [Tanksley et al., 1995]. This strategy begins with attempts to find tightly linked markers to the gene of interest that are located at a physical distance less than the average insert size in the library. The library is then screened with the marker to target a single clone thus eliminating the need for a chromosome walk. Finding linked markers is discussed in a later section.

V: GENOME MAPPING IN SOYBEAN

Soybean (2n=2x=40) is a large (1.1 x 10^9 bp) and complex genome [Blackhall et al., 1991]. Approximately 60% of the soybean genome is repetitive DNA [Goldberg 1978] which is consistent with the repetitive DNA content of higher plants [Zimmermann and Goldberg 1977]. Between 30 to 45% of the genome is organized into long stretches of repetitive sequences which are at least 1500 nucleotides long. Approximately 35 to 50% of the genome is arranged in a short-period interspersion pattern consisting of an alteration of repetitive sequence elements (averaging 200 nucleotides) and single copy DNA. The majority of single copy sequences average less than 3 kbp in length, therefore, a minimum of 250,000 diverse single copy DNA exists in the genome. About 65 to 70% of single copy sequences comprising 1.1 to 1.4 kb alternate with repetitive sequence elements of 0.3 - 0.4 kb [Goldberg 1978, Gurley et al., 1979]. These estimates are based on reassociation kinetics and not direct analysis of the genome. Pachytene analysis has shown that over 35% of the soybean genome is made up of heterochromatin [Singh and Hymowitz 1988]. The isolation of the SB92 satellite sequence substantiates these findings [Kolchinsky and Gresshoff 1995]

Iowa State University soybean map

The first soybean RFLP linkage map was developed at Iowa State University (USDA-ARS/ISU) using an interspecific cross between *G. max* (A81-356022) and a *G. soja* accession (PI 468.916) [Keim et al., 1990]. The development of a saturated soybean RFLP map has been slow for several reasons. In addition to genome size and complexity, soybean is a self-pollinating plant, controlled crosses are difficult to make and large numbers of hybrid seed cannot be obtained. The published map of 1993 [Shoemaker and Olson 1993] contains about 400 polymorphic markers that have been placed in 23 linkage groups representing about 3000 cM [Shoemaker and Olson 1993]. These markers include random genomic clones, cDNA clones, isoenzyme loci, and a variety of morphological and developmental classical markers. In addition to the *nts* locus (see below), RFLP markers

linked to approximately 16 quantitative trait loci including hard seededness [Keim et al., 1990a], seed protein and oil [Diers et al., 1992b] and several genes for seed coat color, actin gene, sharp pubescent tip, acid phosphatase, and vegetative storage protein have been identified [Keim et al., 1990b, Diers et al., 1992b].

University of Utah soybean map

An intraspecific RFLP map was constructed in 1993 at the University of Utah [Lark et al., 1993] in order to utilize the strategy of transgressive segregation. Transgressively segregating progeny are progeny whose expression of a phenotype exceeds that of one or both parents. This type of breeding is important to breeders who want to develop pure line varieties in self-pollinated species [Allen 1994]. In order to breed progeny that have the desired qualities but minimum weaknesses of parental lines, those lines that have different favorable alleles for the trait are selected. Therefore parental lines that are themselves products of pure line varieties that have similar phenotypes but different genotypes must be selected.

The intraspecific map was constructed using F3 progeny from an intraspecific cross between cultivars Minsoy (PI 27.890) and Noir 1 (PI 290.136). The map contains 132 RFLP, isozyme, morphological, and biochemical markers in 31 linkage groups covering 1550 cM of the soybean genome and several important agricultural traits like early maturity, plant height, lodging, seed protein and oil have been measured in this RI (recombinant inbred) mapping population [Lark et al., 1994, Mansur et al., 1993]. Subsequently, the same F3 generation was advanced to an immortal F11 population (recombinant inbred lines) which has since been used as a mapping population in our laboratory. The mapping program applies a correction factor to linkage data from the F11 population before incorporating it into the original data from the F3 segregating population. Therefore linkage data using RILs can be placed on the original F3 map. (Advantages of RI mapping is discussed in a later section)

DuPont soybean map

The DuPont soybean map [Rafalski and Tingey 1993] has over 3000 anonymous low copy soybean genomic clones, forty eight biochemically defined loci, several RAPD markers, and was created using *G. soja* PI81762 and *G. max* cv. Bonus. However, the map is not available for public use.

Soybean map integration

To increase the potential of a molecular map, it is essential to integrate molecular, classical and physical maps. The presence of classical, qualitative markers on a molecular map serves many purposes. Firstly, it makes the map more saturated. Secondly, it provides additional information which can be used for analyzing complex traits. Thirdly, it provides the opportunity to speed the introgression from one cultivar to another, of those characters like flower pigmentation or developmental characters defined by classical markers.

Of the 250 identified markers for pigmentation, morphology and isozymes, only 63 have been assigned to the classical genetic map [Palmer and Hedges 1993]. Most of the 19 classical linkage groups (CLGs) are just 2-or 3- point linkages. Shoemaker and Specht [1994] have initiated a study to combine the linkage groups of the classical linkage map with their homologous equivalents in the soybean RFLP map using a *G. max* x *G. max* cross ('Clark' and 'Harosoy'). Soybean CLGs 1, 2, 3, 4, 7, 10, 14, 18 and 19 have now been associated with a corresponding MLG (Molecular Linkage Group) of the Iowa State map [Shoemaker and Specht 1994].

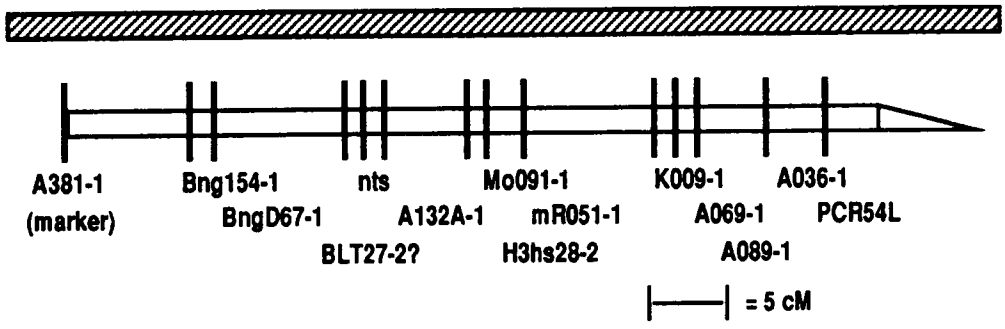
Integration of the classical and molecular map can also be accomplished by using trisomic lines which are now available for soybean [Kollipara et al., 1994] and by using FISH (Fluorescent *In Situ* hybridization). FISH is a technique that allows "painting" of chromosomes with fluorescent probes to localize YACs or other probes with *in situ* PCR [Keim et al., 1995]. Chromosome assignments of a gene can be made based on the segregation of progeny from a cross of nts382 and each trisomic line. Progeny analysis made it possible to assign MLG L to chromosome 13 containing lipoxygenase and the NOR (Nucleolus Organizing Region) [P. M. Gresshoff pers. comm.].

The construction of the first Yeast Artificial Chromosome (YAC) library is underway in the laboratory of Gresshoff [Gresshoff and Landau-Ellis 1994, Funke et al., 1994] and will provide the clones for correlating physical distances with genetic distances in the soybean genome. The maximum insert size in this YAC library is 960 kb and the YACs are reported to be stable. This is an important advantage since there have been reports of YAC instability in terms of deletions during propagation in *Arabidopsis* [Putterill et al., 1993] and humans [Foote et al., 1992].

Characterization and current status of the nts locus

The *nts* mutation is homozygous, recessive, and segregates as a single Mendelian locus [Landau-Ellis et al., 1991]. The *nts* mapping population was derived from a cross between *G. max* line nts382 and *G. soja* (PI 468.397) the wild progenitor. The choice of the parents was based on the fact that a RFLP linkage map of soybean had been constructed [Keim et al., 1990] using *G. max* and *G. soja* and a large number of RFLP probes that had been used to construct that map were available. One of these, pUTG-132a was shown to be tightly linked to the *nts* locus, and has been mapped to linkage group H (Fig. 2) [Landau-Ellis et al., 1991]. The analysis of 171 F2 progeny has shown only one recombination event between the *nts* locus and pUTG-132a. The one recombinant has been used to confirm the marker order as pA-36 - pUTG-132a - *nts*- pA-381. The markers on either side of the pUTG-132a marker are pA-381 (3.4 cM) and pA-36 (12 cM). These distances are different from the Iowa map due to the different parentals used in the Iowa map. Currently, the *nts* region has several additional markers (Fig. 2), Bng154-1 and BngD67-1 from *Phaseolus vulgaris*, mR051-1 and Mo091-1 from mungbean, and H3hs28-2, a histone gene. One of the linkage groups MLGb1, shows a region with a cluster of markers similar to that found on MLGH suggesting duplication of this region of the genome. Since the marker, A132, maps to MLGH (called A132-1) and MLGb1 (called A132-2), BLT27-2 has been putatively assigned to MLGH based on linkage to A132-2 on MLGb1.

The STS/RFLP site defined by pUTG-132a was mapped to two separate alleles of the *nts* locus [Kolchinsky et al., 1994] and provides further support to the previous complementation data [Delves et al., 1988, Landau-Ellis and Gresshoff 1992]. Further support for the genetic data describing *nts* as a single copy gene came from the detection of a single genomic copy detected by probe pUTG-132a. This has now been confirmed by FISH analysis using pUTG-132a as a probe onto soybean chromosomes [Keim et al., unpublished data]. The STS/RFLP may be as close as 0.3 cM of the *nts* locus. It is feasible therefore to use a positional cloning strategy towards cloning the *nts* gene. Since markers on either side of the *nts* region had been mapped, it is possible to start a "chromosome walk" towards the gene. Physical mapping of the pA-36 region (about 10-15 cM from *nts*) by Funke et al., [1993] has shown that the *nts* locus may be less than 150 kb from the pUTG-132a marker and the STS. The other closest flanking marker is pA-381, and this appears to be within less than 4 cM. Thus pA-381 may be too distant to be useful for the isolation of *nts* and further efforts are needed to discover closer markers.



Bng154-1 and BngD67-1 = from *Phaseolus vulgaris*
nts = defined by supernodulation alleles nts382 and nts1007
A132A-1 = defined by RFLP probe pUTG132a as well as PCR markers[b]
PCR54L = STS detected after BSA of homozygous F2 plants[c]
BLT27-2 = putative marker linked on MLGb1 to A132-2.
Assignment possible due to genome duplication[d]
mR051-1 and Mo091-1 = from mungbean map (Young)[e]
H3hs28-2 = unpublished histone gene (Kanazin and Shoemaker[f]

Figure 2: Distal portion[a] of molecular linkage group 'H' showing the supernodulaion gene (*nts*) region in soybean. Approximate distances are given according to mapping done by Dr. R. C. Shoemaker (pers. comm.). Sources of the different markers are given below.

[a] from Dr. P. M. Gresshoff

[b] Landau-Ellis et al., [1991].

[c] Kolchinsky et al., [1995].

[d] from Dr. Ben Mathews, USDA, Beltsville

[e] from Dr. Nevin Young, University of Minnesota

[f] from Dr. Randy Shoemaker, Iowa State University

Several different strategies to find additional linked markers to the *nts* gene and other nodulation genes are being pursued in our laboratory. Bulked segregant analysis, tec-MAAP (Template Endonuclease Cleaved-Multiple Arbitrary Amplicon Profiling), DAF (DNA Amplification Fingerprinting), and SCAR (Sequence Characterized Amplified Region) markers (see below) are being used for finding linked markers to *nts* and these linked markers will be used to screen YACs and contigs containing the *nts* locus.

VI: MOLECULAR MARKERS TO ASSIST IN MAP-BASED CLONING

General properties of molecular markers

The preceding section described the importance of maps for various applications. This section deals with molecular markers used to construct genetic and physical maps. Highly polymorphic behavior as well as frequent, evenly spaced, and random occurrence in the genome, are the most important attributes for a molecular marker. In addition, high reproducibility, an assay procedure that is easy and fast, accessibility of materials for the assay, rapid exchange of data and information between laboratories, and codominant inheritance enhance the utility of a molecular marker. These properties are exploited by the methodologies described below for revealing molecular markers.

Protein markers

Naturally occurring polymorphisms have proved extremely useful as genetic markers. Isozymes in *Drosophila pseudoobscura* is the first example of polymorphic loci demonstrated by Lewontin and his colleagues [Hubby and Lewontin 1966] in natural populations. Isozymes is a collective term that represents multiple molecular forms of enzymes that differ in electrophoretic mobility. Different enzymes that are coded by different loci are referred to as isoenzymes, different alleles of a locus that code for an enzyme are referred to as allozymes, and different post-translational modifications of an enzyme are called secondary isozymes. Differences in isozyme numbers and isoenzyme properties can be used for evolutionary studies, but quantitations of genetic variations among or within populations are obtained only from frequencies of allozymes. In plants, seed storage proteins or enzymes are often used in studies of systematics, population structure and phylogeny. For example, albumins and globulins have been used to infer the 2N progenitors of polyploid species in *Brassica* [Yadava et al., 1979]. Rubisco (Ribulose 1,5-Bisphosphate Carboxylase/Oxygenase) has been used for studying modes of

inheritance from the perspective of systematics and phylogeny [Crawford 1990], and cytochrome c (based on its amino acid sequence) was used for phylogenetic studies among angiosperms [Boulter et al., 1980].

DNA markers

Variations in DNA due to base changes, insertions, or deletions are termed DNA polymorphisms. Important and relevant findings concerning the nature of DNA polymorphisms came from *D. melanogaster* and maize. For example, sequence analysis of the *bronze1* allele in maize [Ralston et al., 1988] showed that in a region of over 3200 bp among three alleles, they detected 115 base pair changes. These changes accounted for 77% of the total of the different types of base pair changes, only 4% being those of major deletions/insertions [Ralston et al., 1988]. The frequency of polymorphism (1.5%) within the coding region was similar to that in non-truncated and flanking regions. Among the eleven wild-type *alcohol dehydrogenase (Adh)* alleles of *D. melanogaster*, Kreitman [1983] detected 43 base pair changes and length polymorphisms (less than 40 bp), but only one of these changes led to a change in an amino acid resulting in a charge altered protein. To summarize, these studies showed that base pair changes are more frequent than large rearrangements, that the level of heterogeneity in coding regions is similar to that found in the rest of the genome, and that polymorphisms in DNA sequences is more frequent than changes in proteins.

VII: METHODOLOGIES TO DETECT DNA SEQUENCE VARIATION

Advantages of DNA-based markers

There are several advantages to using genetic markers based on DNA polymorphisms rather than protein or phenotypic markers: (1) the genotype rather than the phenotype is assayed. There is potential for a considerable loss of information in going from DNA to mRNA to protein. The complete sequence of the gene cannot be inferred from the mRNA or the protein since information contained in the introns is lost in the mRNA and the degeneracy of the genetic code limits the information provided by the protein marker; (2) one or more DNA sequences appropriate to a problem can be selected on the basis of evolutionary rate or mode of inheritance; (3) the methods of analysis can be applied to any DNA; and (4) DNA can be isolated from small amounts of tissue and is relatively stable.

Techniques that detect DNA polymorphisms

The technique of *Restriction Fragment Length Polymorphism (RFLP)* was first proposed by Botstein et al., [1980]. Most frequently used for genetic linkage maps, RFLP technology provides a powerful tool for detecting sequence variation. The technique involves comparison of the number and size of fragments produced by digestion of the genomic DNA with restriction endonucleases. The fragments are visualized by Southern hybridization in which target specific radiolabeled probes are used. The detection of polymorphic fragments (allelic variants) depends on the presence of an altered restriction site or an insertion/deletion in the corresponding allele. RFLP markers have the advantage of being co-dominant and can therefore detect multiple alleles in the region of interest. Since RFLP markers have the ability to detect parental alleles as being distinct, heterozygotes in the F2 progeny can be distinguished from the parental genotypes.

Both cDNA and genomic libraries can be used as RFLP probes. Although more difficult to construct, cDNA probes have the advantage that they map to actual genes. Genomic clones on the other hand encompass all regions of the genome and map to non-coding and more importantly to 5' and 3' flanking regions of genes. The methylation sensitive enzyme *Pst*I is used in plants since it produces small unmethylated sequences that correspond to low copy sequences. Ribosomal DNA probes are conserved and the high copy number makes it easier to detect polymorphisms because of variation in spacer length. These probes have been used very successfully in population biology and systematics [Hamby and Zimmer 1992, Learn and Schaal 1987]. RFLPs in chloroplast and mitochondrial DNA have been used to study organelle transmission, taxonomic and phylogenetic studies at an interspecific and intergenic level [Clegg 1993, Palmer 1985].

Repeated sequences are sets of tandemly repeated oligonucleotide core sequences that have proven to be extremely powerful markers within human populations for both, forensic purposes and diagnostic purposes [Kirby et al., 1990]. *Simple Sequence Repeats (SSRs)*, also known as *microsatellites* or *Short Tandem Repeats (STRs)*, are tandemly repeated DNA sequences that are abundantly present in eukaryotic genomes. SSRs are different from minisatellite repeats (VNTRs) in terms of their core sizes. VNTRs have core sequences that vary in length from 11 to 60 base pairs [Jeffreys et al., 1985a] whereas SSRs have core sequences that are di-, tri-, or tetra-nucleotide repeats. In plants (AT)n sequences are the most abundant followed by (A)n + (T)n, (AG)n + (CT)n, (AAT)n + (ATT)n, (AAC)n + (GTT)n, (AGC)n + (GCT)n, (AAG)n + (CTT)n, (AATT)n + (TTAA)n, (AAAT)n + (ATTT)n, and (AC)n + (GT)n [Wang et al., 1994]. Extensive work done in soybean by Akkaya et al., [1992] showed a high level of SSR length

polymorphism among a diverse group of soybean genotypes with AT being the most common motif (one (AT)n occurring every 250 kbp of genome). A total of 23 SSR alleles was shown to exist at the 17E7 locus among 96 soybean cultivars and germplasm accessions which demonstrates the highly polymorphic nature of this type of marker [Cregan et al., 1994]. In barley, among 71 variants in 207 accessions of wild and cultivated barley, 28 and 37 alleles were found at two SSR loci [Saghai-Maroof et al., 1994]. Due to their highly polymorphic nature and the ability to determine the exact length of the alleles, SSRs have great utility in DNA profiling for plant variety protection. They are also highly suitable for high throughput applications such as assessment of purity in F1 hybrid seed lots. Microsatellites show one disadvantage in that they show unusually high rates of mutation by expansion or deletion [Richards and Sutherland 1992, Mao et al., 1994]. Also, they require a larger input of resources to detect the SSRs (e.g., sequencing). This is not feasible for many species without medical or agronomic importance.

Long, tandemly repeated, sequences are another useful source of molecular markers. They are polymorphic, species or variety-specific, and occur in many cases on centromeres [Laurent et al., 1994] and telomeres. They are also useful for *in situ* hybridization since they are chromosome specific. The soybean 92 base repeat mentioned earlier has been shown to be extremely polymorphic and is organized in giant tandem blocks of over 1 Mb [Kolchinsky and Gresshoff 1995].

Telomeres, defined as the ends of chromosomes, comprise a GT-rich strand that runs in the 5'-3' direction to the end of the chromosome. A consensus sequence, $T/A_{1-4}GG_{1-8}$, is found at the ends of the chromosome that is closely conserved across evolutionary lines [Henderson and Blackburn 1989]. The basic telomeric unit, TTTAGGG in *Arabidopsis thaliana* is only slightly deviated in other plants [Richards and Ausubel 1988]. *Telomeric associated sequences (TAS)* can be used as markers for the ends of chromosomes. They can be used to produce primers that produce highly polymorphic fingerprints. Some polymorphic bands are codominant, and therefore can be easily scored in segregating populations. Lastly they can be cloned and converted into RFLP markers or SCARs (see below) [Kolchinsky et al., 1995]. TAS markers are useful in fingerprints and for defining the ends of chromosomes but less useful for general genomic analysis.

Selective Restriction Fragment Amplification (SRFA) invented by Dr. Mark Zabeau [Zabeau 1993] is used for genome mapping and genotype identification. The

polymorphisms revealed by this technique are referred to as *AFLP** , (*Amplified Fragment Length Polymorphism*). The starting template for SRFA is DNA that has been cut with two restriction enzymes, a rare cutter and a frequent cutter. The two enzymes produce fragments that are ligated to two different adaptors. The design of PCR primers is such that the 3' nucleotides of the primers extend into the unique sequences of each restriction fragment. These nucleotides serve as selective bases to amplify only those restriction fragments with complimentary sequences.

Genomic mismatch scanning (GMS) and *Representation difference analysis (RDA)* are techniques that represent the very latest advancement in genome scanning technology. The idea is to scan the entire genome for polymorphisms as opposed to dependence on finding linkage analysis polymorphisms. Linkage polymorphisms are dependent on detecting only those areas of the genome that have a particular restriction site (RFLP), primer site (DAF, RAPD), or other sequence dependent traits (SSRs, STSs, SCARs, see below). RDA [Lisitsyn et al., 1994] and GMS [Nelson et al., 1993] are techniques that seek to overcome the drawback of determining the sequence of each marker in traditional linkage analysis techniques. Both these techniques use the concept of hybridizing single stranded DNA from two different genomes and select for either mismatched hybrids or those regions of the genome that show identity-by-descent.

Allele-specific PCR markers [Wu et al., 1989] are used to define alleles that are variants in DNA sequence. Specificity is achieved by designing one or both primers so that they partially overlap the site of sequence difference between the amplified alleles. *Single-strand conformational polymorphism (SSCP)* [Orita et al., 1989] takes advantage of the secondary and tertiary structure differences between denatured and rapidly cooled amplified DNA fragments that differ slightly in their DNA sequence. *Denaturing gradient gel electrophoresis (DGGE)* [Myers et al., 1987] resolves partially denatured DNA in precisely defined conditions of temperature and denaturant concentration. Different alleles may denature to various extents under these conditions resulting in polymorphisms. Another approach with some advantages of the RFLP assay but does not need Southern blotting is named *CAPS (Cleaved Amplified Polymorphic Sequence)* [Konieczny and Ausubel 1993]. In this method, partial DNA sequence information from the locus of interest is used

* AFLP is used as a term to describe polymorphisms revealed by DAF or similar techniques. recently, M. Zabeau (Key gene Inc., Wageniyer, The Netherlands) developed an alternative technique (SRFA) which is commercially referred to as AFLP. This thesis uses the term AFLP as first used by Caetano-Anollés et al (1991).

to create a set of primers. DNA from several different individuals is amplified using these primers, and the resulting fragments are digested with a number of restriction fragments to identify RFLPs among individuals .

VIII: ARBITRARY PRIMER TECHNOLOGY (APT)

The discovery of the polymerase chain reaction in 1985 [Saiki et al., 1985, Mullis and Faloona 1987] has transformed the way DNA analysis is carried out in research and clinical laboratories. Developed by Kary Mullis at Cetus Corporation, the technique involves the *in vitro* enzymatic synthesis of millions of copies of a specific DNA segment. Currently, the techniques of DNA Amplification Fingerprinting [Caetano-Anollés et al., 1991b], Random Amplified Polymorphic DNA [Williams et al., 1990] and Arbitrarily Primed - Polymerase Chain Reaction (AP-PCR) [Welsh and McClelland 1990] have been very successful for use as molecular markers. All three techniques although having dissimilarities (see part III), are based on the enzymatic amplification of different DNA segments with a single arbitrary primer to generate a characteristic fingerprint. If two different genomic DNAs are used (e.g. the parents of a cross), the products will be diagnostic for each parent, allowing the detection of polymorphic DNA fragments from the two parents.

The polymorphic bands arise either from a change in the primer binding sites on the template DNA or, from insertions/deletions that change the size of a DNA segment without preventing its amplification or both. The distinguishing characteristics of arbitrary primer technology is that, under relaxed cycling conditions (unlike PCR conditions), the fingerprint is influenced by the sequence, length of the primer, and the chemistry of the reaction conditions. No prior knowledge of the target sequence is required and the primers to bind to several sites on the DNA allowing the entire genome to be scanned for polymorphisms. Other important aspects pertinent to this dissertation are discussed below.

Codominance and the nature of target sites that are amplified

Markers generated by arbitrary methods are generally found to be dominant i.e., the parental genotypes are distinguished by the presence or absence of the amplified marker. However, within limits of the amplification reaction, a change in the size of the amplified marker should result in a codominant marker. A very small percentage of APT marker bands are co-dominant, only 2% of the total number of bands in *L. sativa* [Kesseli et al., 1993] and 0.16% of the bands in *Melamspora lini* of flax rust [Ayliffe et al., 1994] have been shown to be co-dominant. Comparatively, RFLP markers are codominant i.e., both

parental alleles are detected. With APT markers, only two classes of progeny can be distinguished, those progeny that exhibit the presence of the band (homozygous for the presence of band plus heterozygotes) and those that do not exhibit the band (homozygous for the absence of the band). The genomic copy of an amplified sequence cannot be linearly quantitated with the intensity of the marker in the heterozygote i.e., the heterozygote does not show a 50% decrease in intensity compared to the parental band [Prabhu and Gresshoff 1994, and Devos et al., 1992, also see part IV]. It is therefore not possible to estimate the zygosity of a dominant APT marker with existing methods (although the Livak et al., USPTO patent claims to distinguish RAPD heterozygotes from homozygotes).

Amplified fragments have been shown to originate from different types of sequences in the genome. In soybean amplified fragments hybridized to genomic DNA showed that 50% hybridized to low copy DNA, 27% hybridized to middle repetitive and 18% hybridized to highly repetitive DNA [Williams et al., 1993, also part VI]. This suggests that APT markers provide accessibility to all regions of the genome.

Competition in amplification

When genomic DNA from two different individuals is present in the amplification reaction, competition can affect the outcome of amplification. Competition is defined as a change in the ability to amplify a polymorphic band in the presence of DNA from another individual. Studies indicate the amount of DNA contributed by each individual and the type of polymorphic band is important for successful amplification when two individuals from the same species is amplified. For example, in a competition study, two different *Silene alba* individuals, each of which showed a polymorphic band specific to the individual, were chosen for a mixed-template experiment [Williams et al., 1993]. When DNA from the two individuals was mixed in different ratios ranging from 0-100%, polymorphic bands specific to both individuals were amplified if there was less than 80% DNA from any one individual. Another study with individuals from the same species, *Latuca saligna* and *Latuca sativa* (lettuce), showed that the rarer polymorphic band was not detected if it constituted less than 0.2-0.4 of the proportion in the mixture [Michelmore et al., 1991].

When DNA from genomes of different complexity namely *Glycine max* and cyanobacteria were mixed for amplification, even a large excess (460-fold) of cyanobacterial DNA did not amplify the prokaryotic sequences when a 10 nucleotide primer was used. It was suggested therefore that genome complexity was important presuming that more complex eukaryotic

genomes could provide a larger selection of primer sites. It was predicted that shorter primer sites (7.5 mers) would produce a higher proportion of prokaryotic amplification bands [Williams et al., 1993]. Experiments using *Azolla* and *Anabena* symbiotic systems [Eskew et al., 1992] showed that although shorter primers (8 mers) did produce more prokaryotic amplified products, the sequence of the primer was more important in determining which DNA was preferentially amplified.

Several lines of evidence suggested that genome complexity cannot be regarded as the sole explanation for preferential amplification of eukaryotic DNA from a mixed template. First, the contribution of *Anabena* sequences ranged from 0-77% of the total number of amplified bands when a mixture (intact symbiosis) was used for amplification [Eskew et al., 1992]. Second, experiments by Caetano-Anollés et al., [1992a] show that the 8 bases at the 3' end of the primer are critical for a given fingerprint (see below) and that extra bases at the 5' end as in a 10 mer, have a limited effect in the profile. Third, the use of excess primer as in the DAF reaction should theoretically bind to all available sites in the genome.

Primer length and sequence

Primer lengths from 5 up to 35 bases have been used in arbitrary primer-generated DNA fingerprints. Primer sequence, length and concentration all affect the outcome of the amplification reaction and depending on the detection system [Caetano-Anollés et al., 1992a and part III] also affect the complexity of the fingerprints obtained. Sommer and Tautz [1989] showed that primers (specific as for PCR) between 17 and 20 nucleotides require at least 3 homologous nucleotides at their 3' end for successful priming. Weak priming was also obtained with a long primer having only two nucleotide homology with the template. Information about overall effects of primer length and sequence used in arbitrary primer amplification came from studies by Caetano-Anollés et al., [1992a]. In a series of elegant experiments they showed that primer sequence and length was an important parameter in the specificity of the DAF reaction. They showed that for arbitrary primers, eight nucleotides was required to produce stable and repeatable profiles (possibly by avoiding primer-template mismatching). Amplification profiles with longer primers were identical or related to those produced by 8 mers whereas profiles with primers shorter than eight nucleotides gave completely different patterns. The 3' end of the primer sequence was more sensitive to single base changes than more distal changes. The 8 base boundary specified by the primer was not changed by increasing the annealing temperature. Experiments using primers of increasing length derived from a common sequence showed

that not all primed template sites are amplified. Factors other than primer-template homology play a part in the selective amplification of some products [Caetano-Anollés et al., 1992a].

Based on these results, a general hypothesis explaining the importance of primer sequence and length was proposed by Caetano-Anollés et al., [1992a]. In this model, DNA amplification is controlled at two levels by different factors. In the first few cycles of amplification the primer screens potential template sites and a subset of a large number of possible target sites is selected. Factors affecting this initial selection are components of the reaction including cycling parameters. In the second stage, effective amplification of the selected products of the first round depends on the internal sequence of the product. Any sequences that produce hairpin like structures are likely to be inefficiently amplified and the extent of this hairpin symmetry allows some products to be efficiently amplified.

Tailoring fingerprints

Fingerprinting with pairs of arbitrary primers referred to as "multiplex fingerprinting" generates a pattern different from that of the individual primers thus increasing the repertoire of fingerprints that can be generated [Caetano-Anollés et al., 1991b, Welsh and McClelland 1991b]. The extent of the difference between the multiplex pattern and the individual patterns depends on the template and the primer. This technique has been used very effectively in soybean [Caetano-Anollés et al., 1991b] and recombinant inbred mice [Welsh and McClelland 1991b] and this method has been reported to increase detection of polymorphisms in turfgrass and rat [Callahan et al., 1993; Micheli et al., 1993]. Mini-hairpin primers contain a core of arbitrary sequence and a stable mini-hairpin at the 5' end. An increased number of products was observed with mini-hairpins in near-isogenic lines of soybean, *G. max* cv. Bragg and line nod49 (S. Abbitt, pers. comm.). Mini-hairpin primers have been successfully used in analysis of sub-genomic molecules like YACs, cloned DNA and PCR products [Caetano-Anollés and Gresshoff 1994].

tecMAAP (Template Endonuclease Cleaved Multiple Arbitrary Amplicon Profiling)

tecMAAP is a fingerprinting technique that was designed to increase the power of APT. If genomic DNA is first digested with a restriction endonuclease (e.g. four base cutter) and then subjected to DAF, the probability of obtaining polymorphic fragments was increased. Tolerance for primer-template mismatches and efficient amplification of some of the amplicons are possible explanations for this enhanced detection [Caetano-Anollés et al.,

1994a]. By this method, 42 polymorphic bands tightly linked to the *nts* locus were detected between near-isogenic lines, *G. max* cv Bragg and line nts382 a mutant line derived from Bragg. The power of this technique for distinguishing near-isogenic lines is illustrated by the fact that without pre-digestion, the same number of primers were unable to detect polymorphisms [Caetano-Anollés et al., 1993].

Use of APT markers for mapping

APT markers in plant genome mapping

High resolution maps impact three major areas of plant breeding: understanding gene action, transfer of desirable genes through genome manipulation and "marker-assisted selection". Traditionally, breeders relied on morphological and other phenotypic markers for selection in breeding programs. Although phenotypic markers together with sophisticated statistical methods have been successfully used for selection, some of these markers have the disadvantage of varying with the environment e.g., a resistance trait may not be expressed because of dry conditions. The development of molecular markers therefore represented a significant improvement as they offered greater diversity and could be detected independent of the environment.

Most of the linked markers to QTLs (Quantitative Trait Loci) in major agronomic crops have been RFLP markers. For example, in soybean, a number of QTLs for lodging, days to flowering, days to maturity, leaf length, leaf width, yield, seed weight, seed oil and seed protein have been identified by establishing linkage to RFLP markers in the recombinant inbred soybean population [Orf et al., 1994]. In maize, however, QTLs involved in the resistance to the pathogen *Gibberella zeae*, which causes stalk and ear rot were mapped by combining RFLP and RAPD analyses of F2 individuals [Pé et al., 1993, also part IV]. Markers linked to qualitative traits e.g., the *Phytophthora* resistance loci in soybean [Diers et al., 1992a], a nematode resistance gene in tomato [Klein-Lankhorst 1991], and the *nts* locus in soybean [Landau-Ellis et al., 1991] have also been accomplished due to the availability of RFLP markers. In soybean, RAPD markers linked to traits genes controlling sudden death syndrome (SDS) [Hnetkovsky et al., 1994], frog-eye leaf spot [Qui et al., 1994], soybean cyst nematode (SCN) resistance [Skorupska et al., 1994], and in sugar beet, RAPD markers linked to a nematode resistance gene and a gene for hypocotyl color [Uphoff and Wricke 1992] have been reported. Due to their ease of detection, APT markers linked to QTLs (see part IV) will be extremely useful for marker-assisted selection in

breeding programs by facilitating the screening of a large number of progeny.

Mapping populations for APT markers

The choice of a mapping population depends on the breeding system of the plant. Most RFLP maps have been produced using F2 populations or backcross populations derived from crosses between inbred parent lines. Carlson et al., [1991] have determined that using an identical linkage value of p = 0.2, 280 individuals would have to be analyzed to obtain the same amount of information among dominant/recessive class of progeny (as in DAF) compared to 50 for codominant class of progeny (as in RFLP). The problems of distinguishing heterozygotes and/or the need to increase F2 sample size can be circumvented by using a mapping population of recombinant inbred lines. Using inbred parents simplifies genetic analysis because the phase (coupling or repulsion) of the marker is known. The combination of alleles defines coupling (e.g., AA or aa) and repulsion (Aa). Since two recombinant chromosomes can be scored in each plant, F2 populations provide more mapping information with codominant markers. F2 populations also provide a sex averaged map because chromosomes from both the male and female are scored in the progeny. Backcross populations can provide a male or female map depending on which sex was the recurrent parent. The database of markers accumulates as long as the same set of F2 or backcross plants is used. If the mapping population is lost or used up, the mapped markers have to be re-mapped in the new population. Permanent mapping populations prevent a wasteful use of time and resources. F2 mapping populations are less suitable for mapping with APT markers since they are dominant markers. Alternative mapping populations like backcross, recombinant inbred, double haploids do not suffer a loss of mapping information (they are equally informative) with either codominant or dominant-recessive markers.

Those plants (*Solanum chacoense* [Rivard et al., 1989]) in which doubled haploids can be produced by anther culture of an F1, produce a one step homozygous mapping population similar to recombinant lines (see below). Chromosome elimination which occurs in some interspecific or inter-generic crosses can be used to produce homozygous populations by doubling the resultant haploids [Kasha and Sadasivaiah 1971, Heun et al., 1991]. These methods have the disadvantage that changes of RFLP patterns have been observed in tissue cultures [Roth et al., 1989] but these populations are suitable for APT mapping.

Mapping populations that breed true lines (recombinant inbred lines) can be constructed in

selfing plants by single seed descent of individual F2 plants that have been selfed or by single seed descent of backcross plants. After the sixth generation of selfing, the population is considered to be *largely* homozygous (see below) and can be propagated by seeds. In theory, each chromosome of a recombinant inbred line contains loci derived from each mapping parent that are homozygous, and will not recombine, i.e., the recombination events and therefore the mapping distances are fixed in time. Markers that are scored using recombinant inbred lines therefore will be homozygous for each parent. In practice, some regions do tend to stay heterozygous longer than predicted by theoretical calculations [Burr and Burr 1991]. The recombinant inbred population from Utah has approximately 0.04% heterozygotes in the mapping population [Dr. G. Lark, pers. comm.].

In addition to scoring only homozygote parental genotypes, there are several advantages to using recombinant inbred lines (RIL) for mapping purposes. The source of DNA is immortal so there is no upper limit to the number of markers that can be put on the map. Other research groups using the RILs can provide additional markers that can be linked directly to the existing map. Also because seeds can be easily transported to other laboratories, the problem of transferring plant cuttings across and state and international boundaries is avoided. Recombinant inbred maps have higher resolution due to the large number of recombination events that occur in each generation prior to achieving homozygosity. Comparatively, F2 populations have lower resolution since the recombination events in only one generation are being scored. Due to the larger population, any errors in scoring segregation data have a lesser impact in the recombinant inbred population than in the F2 population. For example, an error of ± 2 cM in the F2 linkage data reduces to error to ± 0. 5 cM with the recombinant inbred data. There is about twice the number of recombination events compared to that found in F2 progeny allowing for finer resolution of linked markers.

APT markers are particularly useful for mapping in gymnosperms since they have an elaborated gametophyte stage. Gametophyte tissue is haploid and each female gametophyte produced on an F1 hybrid is derived from a single meiotic event. A set of gametophytes which constituted the mapping population was derived from a single F1 plant to construct a genetic map in conifers [Tulsieram et al., 1992]. The long generation time of forest trees and their large size makes generation of a mapping population rather formidable. Conifers are very suitable for mapping with arbitrary primers because segregating progeny often represent a backcross situation between a homozygote and a heterozygote and the level of heterozygosity among individuals is very high. Using F1 progeny, Carlson et al., [1991]

constructed RAPD maps for Douglas fir and white spruce.

Using APT to find markers linked to genes of interest

The use of "marker assisted" selection for traits such as pest resistance has been greatly limited primarily due to the low number of morphological markers that are available in most crop species. Added to this is the inability to score multiple morphological mutant traits in a single segregating population and their deleterious effects on plant phenotype. The predominant factor in the choice of parents for developing mapping populations in most crop species has been the presence of maximum RFLP variation. Less emphasis has been placed on choosing parents that show high levels of polymorphism for traits of agronomic importance. For this reason, although RFLP maps have been constituted for most agronomically important species, the labor intensive nature of the technique does not generally offer the opportunity to efficiently target tight linkages to specific loci of interest. Two major strategies that have been successfully used for finding tightly linked specific markers, bulked segregant analysis (BSA) [Michelmore et al., 1991] and near-isogenic lines (NILs) [Martin et al., 1991, Haley 1994]. The major advantage of the former method is that it does not require the use of NILs since the production of NILs requires time and resources which may not always be available to researchers.

BSA increases the efficiency of linkage detection with molecular markers by scoring polymorphisms between two DNA samples formed from pools instead of individuals. In classical linkage genetic mapping, the position of a target locus with respect to a marker is determined by analyzing individuals of a segregating population. In bulked segregant analysis, a genetic interval (window) between two mapped markers containing the locus of interest is selected. Segregating individuals (as in F2 progeny) consisting of *AA*, *Aa*, and *aa* genotypes are pooled into two groups. The two groups are *AA* and/or *Aa* (group 1) and *aa* (group 2). The selection of the genotypes can either be based on phenotype or on RFLP banding patterns. If the two groups are now screened with arbitrary primers any polymorphisms that are detected between the two groups of pooled F2 DNA, should show linkage to the locus within the assigned interval. Any polymorphism that is detected in a region outside the gene of interest is nullified by the presence of other F2 individuals in the pool. The number of individuals in each pool can also affect the tightness of the linkage. The advantage of BSA is that NILs, like DNA preparations, are created *in vitro* by bulking the F2 progeny into two pools.

A model developed for phenotypic pooling predicts the window and optimal pool sizes for detecting tightly linked markers [Williams et al., 1993]. The mathematical model predicts that at a density of 1 RAPD/cM, a 10 cM window could be detected by 10 tested RAPD markers. Consistent with the prediction of the model 3 out of 5 mutations were mapped to a region, where markers define a 8 cM window. The model was tested in the genome analysis of *Arabidopsis* in which the density of RAPD markers is 1 RAPD marker/3.9 cM.

Near-isogenic lines are derived by repeated back crossing of the progeny to the parent line until the progeny is genotypically identical to the recurring parent except for the desired trait. The chief use of repeated backcross is to transfer a particular gene from one strain with an undesirable background to another strain with the desired background. The reason why backcross - derived NILs have been successfully used for finding tightly linked markers is that during this type of breeding, the heterozygosity at the desired locus is maintained while homozygosity in other regions is systematically increased. For example using NILs and BSA tightly linked RAPD markers have been found in *Phaseolus vulgaris* L. to the rust (*Uromyces appendiculatus* (Pers.) Unger var. *appendiculatus*) resistance gene [Haley et al., 1993,1994], BSA was used to obtain a fertility restorer gene for the *Ogura* radish cytoplasmic sterility of rapeseed [Delourme et al., 1994], to obtain markers linked to *Lr9* leaf rust resistance genes on wheat [Schachermayer et al., 1994], and to obtain markers linked to the downy mildew resistance gene, *Dm*, in lettuce [Paran and Michelmore 1991]. BSA has been successfully used to detect a polymorphic marker (converted into a SCAR, see below) linked to the cluster of markers in the *nts* region [Kolchinsky et al., 1995]

Future trends in methodologies for APT markers

Sequence Characterized Amplified Regions (SCARs) and Sequence Tagged Sites (STSs)

A SCAR is a genomic DNA fragment at a single genetically defined locus that is identified by PCR amplification using a pair of specific PCR primers [Paran and Michelmore 1993]. An STS is characterized by a short, single copy DNA sequence that can be amplified by PCR from a genomic library or genomic DNA using specific oligonucleotide primers [Olson et al., 1989]. STS information can be used by different labs, independent of the method used in physical mapping thus circumventing the need to store and distribute large numbers of clones. SCARs and STSs are both used as physical landmarks in the genome, however, SCARs are distinct (from STSs) in that they are primarily used as genetic

markers. SCARs can contain both single and repetitive sequences, whereas STSs by definition contain only single copy sequences. The uniqueness of SCARs is determined by the primer sequence and their spacing rather than by hybridization as in STSs. For example, a STS generated from pUTG-132a linked to the *nts* region did not show any polymorphism between several commercial varieties within the sub-species *G. max* suggesting that the sequences contained in the PCR fragment showed a high degree of conservation [Kolchinsky et al., 1995]. SCARs are obtained by cloning and sequencing the two ends of the amplified products of DAF or RAPD markers. The advantages of SCARs are that, they detect only a single locus, their amplification is less sensitive to reaction conditions (unlike RAPDs), and they can be potentially converted into codominant markers.

A large number of STS primers have been produced in human and mammalian systems by analyzing RFLP markers [Tang et al., 1992], microsatellites [Dietrich et al., 1992], *Alu* elements [Nelson et al., 1989], expressed sequences [Durkin et al., 1992], end fragments of YACs [Kere et al., 1992] and end sequences of cosmid inserts [Miwa et al., 1993]. STSs facilitate screening YAC libraries and the construction of long contigs of physical mapping [Chumakov et al., 1992]. STS or SCAR primers have been designed from RFLP markers and BSA in soybean [Kolchinsky et al., 1994] and rice [Tanksley et al., 1992, Williams et al., 1991]. In rice, STS primers have also been synthesized from microsatellites [Zhao and Kochert 1992] and random genomic clones in order to obtain even distribution in the genome [Inoue et al., 1994]. RAPDs have been used to produce SCAR primers in lettuce [Paran and Michelmore 1993] and in barley, published sequence data was used to design SCAR primers [Tragoonrung et al., 1992] (see part VI).

Automation

A major rate-limiting step in all the methodologies described above is the separation and detection of DNA or protein markers using agarose or polyacrylamide gel electrophoresis. When large numbers of objective and accurate DNA fragment length measurements are required, automation becomes desirable. Capillary gel electrophoresis (CGE) and use of fluorescent dyes for labeling DNA are two techniques that show potential for automation.

CGE allows fast separation, needs only minute quantities of sample, and gives high quality reproducibility [Landers 1993]. The basic concept of CE is the use of a capillary with internal diameters ranging from 20-100 μm. This allows a very efficient dissipation of heat

(high surface to volume ratio), as a result of which separations can be performed at voltages up to 30,000 V, and reducing separation time to minutes. Two types of sieving matrices or gels are used for DNA separations, 'chemical gels' which are immobile and covalently crosslinked (polyacrylamide) and 'physical gels' which are mobile, polymeric and non-crosslinked (polyacrylamide or methyl cellulose). The volume of the capillary is typically 1.2 µl, volume of the injected sample can be as low as 0.2 nl, and the application of pressure or vacuum for sample application and rinsing the column between samples decreases separation time to minutes. Combining CGE with Light Induced Fluorescence (LIF) detection of fluorescein labeled DNA, as little as 10^{-20} moles can be detected [Cohen et al., 1990]. Physical gels are used to detect double stranded DNA found in amplification products or restriction fragments, and intercalating dyes like ethidium bromide or cyanine are used for detection. LIF detection using cyanine produces a 1000 fold increase in sensitivity and allows for detection at the attomole level (10^{-19} mole). If polyacrylamide gels containing urea are used, single base separation of DNA is possible. Currently, separation of fragments up to 23 kb [Strege et al., 1991] and intact chromosomes from 3 to 5 Mb [Guszczynski and Chrambach 1991] have been reported.

In the automated fluorescent-based separation technique [Mayrand et al., 1992] DNA molecules are labeled with up to four fluorescent dyes that are different in color, and analyzed within a single electrophoretic lane. The DNA sample is combined with an aliquot of an internal lane standard and electrophoresed. The standard is used to create a calibration curve within each gel lane which is then used to determine the length of a PCR product by comparison. Both the horizontal (agarose) and vertical (acrylamide) gel systems use a detector that is an argon laser beam that scans across the width of an electrophoresis gel at a fixed distance from the sample loading well. The fluorescence that is produced by migrating bands is detected, digitized, and transmitted to a computer for data storage and analysis. A 2 bp resolution was obtained when sizing alleles at three VNTR loci from five related humans [Mayrand et al., 1992].

Other uses of APT in plant genomes

APT markers can also be used to identify new genotypes for increasing genetic variation, to evaluate the genomic composition of each cultivar and to determine the actual genetic contribution of each progenitor to a cultivar. Among other factors, the reproductive method used by the plant influences the level of polymorphism within a plant species. For example, selfing, outcrossing, vegetative propagation modes of reproduction can produce varying

levels of genetic diversity.

With respect to cultivated plant species, RAPDs have been used for cultivar identification of vegetatively propagated apple, blackberries, raspberries, carnation, chrysanthemum, and seed propagated cultivars like cocoa, *Brassica*, papaya, sugarbeet, celery, tomato, allium, cereals and legumes [references can be found in Weising et al., 1995]. A few examples illustrating the discriminatory power of APT markers is given below. In *Stylosanthes* a forage legume, 200 RAPD bands revealed by 22 primers were used to distinguish 4 species. The level of polymorphism was low (0-16%) within species, higher (28-46%) among species, but very low (0-2%) among individuals within each cultivar. The genetic relatedness was in good agreement with morphological, cytological, and isoenzyme studies [Kazan et al., 1993]. The use of arbitrary primer technologies has had widespread use with the economically important cereals, e.g., rice. Using the M13 universal primer, AP-PCR, showed a high number of polymorphisms between rice varieties ssp. indica IR54, japonica and cv. Lemont [Welsh and McClelland 1990]. In a study with RAPD primers, 28 out of 36 primers yielded 116 markers to distinguish 16 rice accessions representing 3 species in *Oryza sativa, Indica, Japonica and Javanica.* Cluster analysis showed *Indica* accessions separate from the two *Javanica* accessions which were closer to *Japonica* than *Indica* which was in keeping with known pedigree relationships [Fukuoka et al., 1992]. In 13 cultivars of wheat, low levels of diversity had been observed by RFLP and RAPD analysis on agarose gels. These estimates of diversity were found to be much higher using DGGE electrophoresis which revealed a high number of polymorphisms [Dweikat et al., 1993]. In another study using 14 cultivars of soft red winter wheat and hard red winter wheat, 5 RAPD primers and 3 combination of two primers revealed a 85% similarity between cultivars and divided the wheat into two main groups, hard and soft, and two subgroups within the soft group [He et al., 1992].

There are several considerations when using APT-based data for genetic relationship studies. (1) The genotype status (hetero- or homozygous) of a band produced by APT is not known. (2) Allelic fragments cannot be calculated since allelic pairs cannot be assigned. (3) Co-migrating bands may not represent homologous loci. (4) Individual bands may not represent independent characters, they may be linked or allelic. Therefore the same band may be counted twice for similarity measurements.

IX: EVALUATION OF MOLECULAR MARKERS FOR PLANT BREEDING AND GENETIC MAPPING

There are some important considerations in evaluating the use of molecular markers for effective genetic diagnosis in plant breeding. The techniques for marker identification linked to the trait of interest must be efficient and can only be achieved through automation. High-throughput technology for DNA extraction and the assay for marker identification is required so that thousands of individuals can be screened. Lastly, user-friendly software for data storage and analysis to enable quick breeding decisions.

In comparison to morphological markers, RFLP markers offer significant advantages with respect to increased numbers of loci and alleles detectable, overall phenotypic neutrality, and the ability to score them at any stage of plant development. In addition, the fragments come from homologous segments although there may be difficulty in distinguishing orthologous from paralogous fragments. The codominant nature of the fragments makes it particularly suitable for QTL analysis and application of statistical methods to study gene flow, population differentiation, and genetic distances. RFLP markers are commonly detected within and among species, which allows the use of RFLP maps of one species to be used in another. However, RFLP-based genetic maps have certain limitations. In crops with a narrow genetic base, like soybean, wheat, and tomato there are few allelic variants of RFLP loci. The RFLP maps constructed for these species have used predominantly inter-specific crosses rather than within cultivars in order to increase the level of genetic polymorphisms. Such maps have limited utility in crosses due to the difficulty in identifying informative (low copy) RFLP markers within such species. Because most traits of agricultural importance are of a polygenic nature (quantitative), a large number of individuals have to be screened to minimize sample variation. Agricultural application of RFLPs therefore incurs increased cost in radioactivity or similar detection reagents and generation of probes.

Single primer methods require nanogram amounts of DNA per primer compared to RFLP analysis which requires microgram amounts of DNA per enzyme tested. In addition, Southern blot analysis is complex therefore RFLP technology is not as amenable to automation. APT markers on the other hand, are generated by a simple, rapid and relatively inexpensive technique making them very suitable for automation. The simultaneous use of two arbitrary primers (multiplex analysis) to generate a different fingerprint than either primer alone, contributes to the cost effectiveness of single, arbitrary primer techniques.

Also, multiplexing increases the number of data points obtained per analysis. APT primers bind at arbitrary sites in the genome and can detect polymorphisms in regions not accessible to RFLP markers particularly repetitive regions. A significant percentage of the probes that are generated for RFLP analysis are not useful because they hybridize to repetitive DNA in the genome.

A good measure of informativeness is expected heterozygosity (H) which is defined as the probability that 2 alleles taken at random from a population are different [Rafalski et al., 1994] ($H = 1 - \sum P_i^2$, where P = allele frequency of i^{th} allele). The higher the H value, the higher is the probability that a particular marker will reveal a polymorphism. Multiplex ratio defines the number of loci that are simultaneously analyzed per experiment, i.e., the number of bands seen on a gel for techniques like DAF, AFLP and RAPD. In case of SSRP markers the multiplex ratio is generally equal to 1. In theory, the best marker systems for mapping depends on both heterozygosity and multiplex ratio. In practice, technical difficulty, cost and ease of performance will play as big a part in determining the utility of the marker systems. Twelve soybean cultivars and PIs [A.K. Harrow, Bonus, CNS, Manchu, Mandarin, Mukdin, Richland, Roanoke, Tokyo, PI54.610, PI81.762 (*G. soja*) and PI 440.913B (*G. soja*)] representing 92% of alleles present in soybean germplasm were used in a comparative study of RFLP, RAPD, AFLP, and SSRP markers. Results indicated that SSRP markers exhibited the highest expected heterozygosity amongst 110 RFLP markers, 37 RAPD bands, 136 AFLP bands and 36 SSRP loci and AFLPs (Key-PCR) showed the least heterozygosity but a very high effective multiplex ratio [Rafalski et al., 1994].

X: Thesis Rationale and Objectives

The feasibility of two different strategies, marker-assisted positional cloning and reverse genetics for cloning the *nts* gene is explored in this dissertation. Since positional cloning necessitates a region saturated with molecular markers the aim of this thesis is to investigate the suitability of arbitrary primer technology for developing markers in the soybean genome. APT, is a recently developed technology that utilizes the polymerase chain reaction. The utility of DAF for generating polymorphisms has already been established [Caetano-Anollés et al., 1991b], hence the aim of this thesis is to expand the scope and utility of DAF by demonstrating inheritance, mapping capability and examining the genomic basis of these polymorphic markers with the ultimate aim of finding markers linked to the *nts* locus. This dissertation is organized into parts and a short synopsis of each

part is given below.

Part II shows the analysis of soybean leaf proteins in order to identify any gene product of the *nts* gene. Two dimensional electrophoresis was used to determine protein differences between wild type and the mutant nts382. This part is included in the thesis since the findings from this study necessitated a change in strategy for cloning the *nts* gene. Part III through part VI deals with the use of DAF for finding markers in the soybean genome. The parts are organized to investigate the suitability of DAF as a technique to reveal polymorphic markers, their mode of inheritance, their usefulness, and their genomic basis. In part III, important aspects of the DAF reaction such as optimization and reproducibility are investigated. Also the silver staining technique is examined for linearity and resolution. In part IV, the inheritance of DAF markers is investigated and DAF markers were mapped in the soybean genome. Part V examines the discriminatory power of DAF markers for applications such as distinction of cultivars and pedigree assessment. In part VI, the conversion of DAF markers into RFLP probes is shown, thus expanding the scope of DAF. The converted RFLP probes are then used as a first attempt to analyse the origin of arbitrary primer markers.

PART II*

POLYPEPTIDE PATTERNS IN LEAVES OF *GLYCINE MAX* (L.) MERR. CV. BRAGG AND ITS SUPERNODULATING MUTANT NTS382 DURING EARLY NODULATION

* Sayavedra-Soto LA, Angermüller SA, Prabhu RR, Gresshoff, PM: Polypeptide patterns in leaves of *Glycine max* (L.) Merr. cv. Bragg and its supernodulating mutant during early nodulation. Physiol Mol Biol Plants 1: 27-36 (1995)

I: INTRODUCTION

In soybean, regulation of nodulation involves shoot-controlled systemic suppression, autoregulation, nitrate inhibition, and arrest of stage III/IV nodule foci. All four phenomena share at least one step that is controlled by the *nts* locus. The model for autoregulation of nodulation in soybean originally proposed by Gresshoff and Delves [1986] is shown in Fig. 1 (for details see part I) [Caetano-Anollés et al., 1990]. The signal 'Q' originating from the root induces the production and/or the systemic release of SDI (shoot-derived inhibitor). SDI arrests further cortical cell division and suppresses nodulation in younger tissues. "X" indicates other systemic signals that may act as positive feedback regulators of nodulation, e.g., in the formation of the first nodules. The supernodulating mutant nts382 is defective at a step subsequent to the production of the signal 'Q'.

The purpose of this study was to determine whether infection of roots of legume plants such as soybean was associated with concomitant polypeptide changes in leaves. A comparison of shoot proteins associated with inoculation in wild type and mutant could potentially detect a candidate protein that represents an altered primary or secondary product of the gene *nts382*. If such a protein that shows altered levels in mutant compared to wild type is found, a probe for isolating the gene could be synthesized by microsequencing the protein spot on the gel.

In this study, proteins from two different sources; total *in vivo* proteins and *in vivo* ^{35}S labeled proteins from inoculated or asymbiotically grown soybean plants were extracted at different time points after inoculation and analyzed by two dimensional electrophoresis. A third source namely, *in vitro* translated products using total RNA, was also analyzed by two-dimensional electrophoresis as part of this study in collaboration with Dr. L. Sayavedra-Soto and S. Angermüller. The results from the analysis of *in vitro* and *in vivo* protein differences is published in Sayavedra-Soto et al., [1995]. It was hoped that this strategy would detect any changes associated with (1) polypeptides controlled at the level of transcription, processing, stability, or transport of their mRNA, (2) total protein reflecting the cumulative effect of all existing proteins at that time point, and (3) *de novo* proteins and those proteins regulated at the translational level. The three techniques would reveal any changes in existing proteins and/or newly synthesized proteins in shoot, associated with autoregulation.

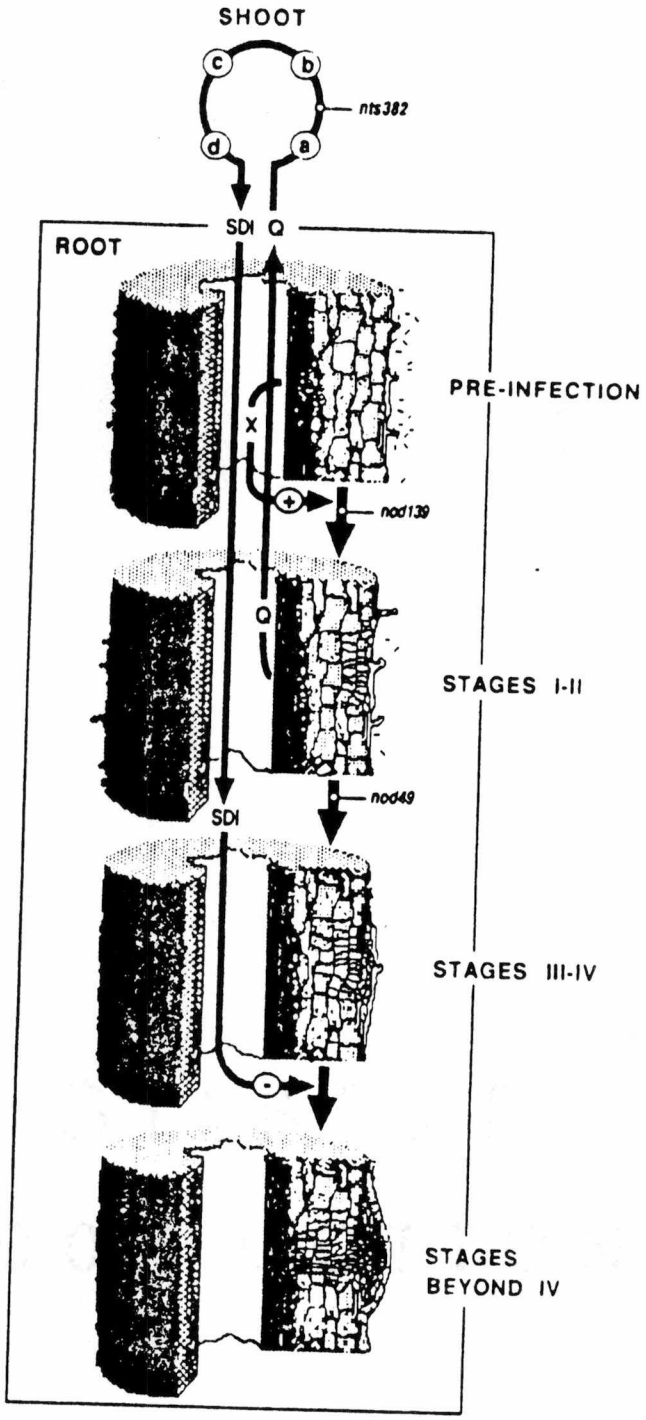


Figure 1: General model for autoregulation in soybean[a]. Shown is a diagrammed root depicting different stages of early events in nodulation. See text for explanantions. SDI, shoot derived inhibitor; Q, signal from root; X, other systemic signals derived from shoot; a-c, unknown precursors of SDI; +, positive activators; -, suppressors.

[a] Taken with permission from Caetano-Anollés, G. and Gresshoff, P. M. (1990). Early induction of feedback regulatory responses governing nodulation in soybean. Plant Science 71: 169-181

II: MATERIALS AND METHODS

The experiments were designed assuming that the shoots of *Glycine max* are responsible for the regulation of nodule number and that the nts mutant is not actively producing the functional factor(s) (and possibly proteins) controlling nodulation. The growth conditions were standardized to account for any differences observed in the polypeptide changes due to inoculation.

The plants grew with no signs of stress in vermiculite itself and showed only slight Mn toxicity after 30 days if mixed with autoclaved sand. Leaf material was harvested only from green, healthy, homogeneous plants. The first visual evidence of nodulation was observed approximately 7 days after inoculation with fully emerged nodules by 10 days. Temporal profiles of nodulation were very similar in 6 different experiments. Sampling times were influenced by key developmental steps. At day 2 cell divisions occur and by day 10 nodules have appeared and nitrogen fixation starts. Later time points represent fully functional nodules (Fig. 2). Low nitrate added to eliminate starvation symptoms in long range cultures did not affect protein patterns of leaves (data not shown). Shoot samples included trifoliolates, primary leaves, petioles, and apices. Control leaf tissue from uninoculated plants was harvested in parallel. Leaf tissue was harvested only once per plant allowing them to grow until the end of the experiment to be checked for nodulation.

Based on the profile of autoregulation described above, leaf tissue was collected at time intervals of 1, 2, 8, 10, and 25 days after inoculation. The leaf tissue was then processed either for protein isolation or RNA isolation. Methods for protein isolation, *in vivo* labeling are described in part VIII, Materials and Methods. All protein samples were then processed for two-dimensional analysis as described in part VIII. RNA isolation and *in vitro* translation procedures can be found in Sayavedra-Soto et al., [1995].

III: RESULTS

Comparison of total protein and ^{35}S labeled in vivo proteins from wild-type Bragg and nts382

Procedures used for total protein extraction, *in vivo* labeling, and two dimensional electrophoresis are described in Materials and Methods (part VIII). Comparison of proteins on the two-dimensional gels (mini and Protean II) was done on a visual basis only.

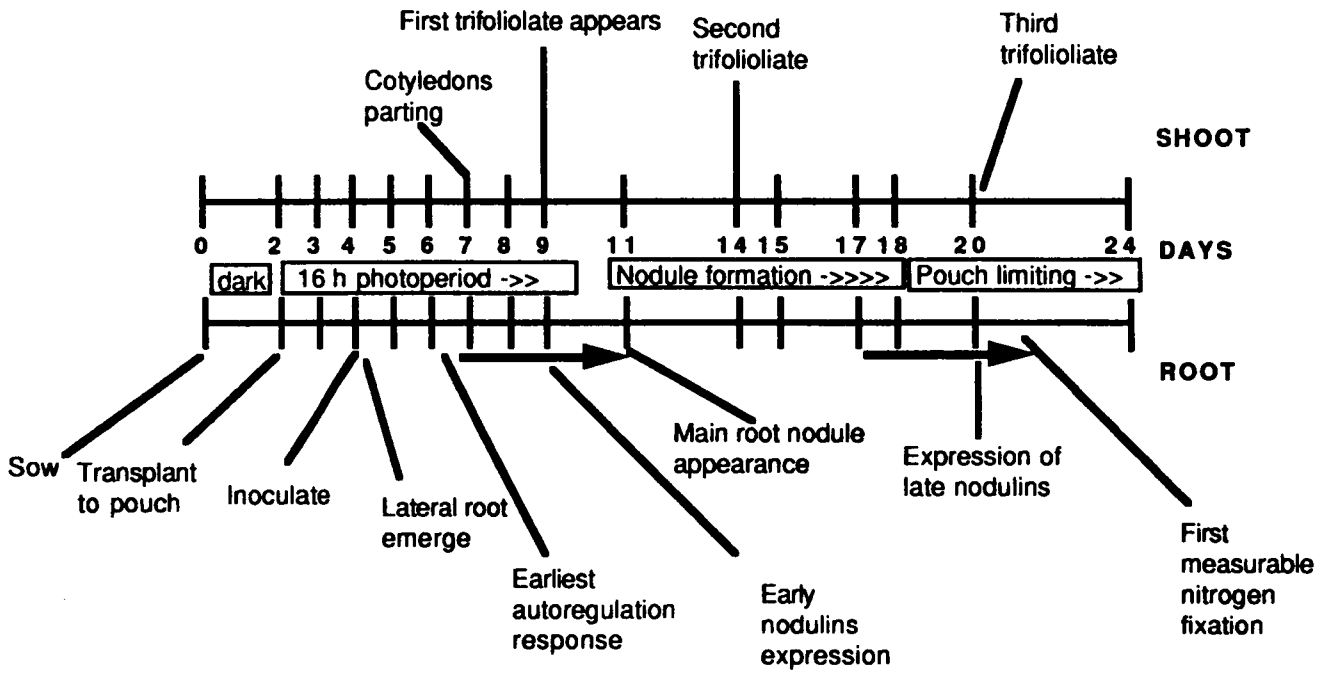


Figure 2: Schematic showing time course of nodulation and autoregulation at different developmental stages of soybean plant growth. Data relate to a growth pouch irrigated with Herridge solution (see Materials and Methods) and incubated at 16/8 hr at 22•C. Biological material is cv.Bragg and *Bradyrhizobium japonicum* USDA110.
Figure kindly provided by Dr. G. Caetano-Anollés.

A comparison of total protein between Bragg (panel A) and nts382 (panel B) detected approximately 200 polypeptides of different intensities ranging from pI 4.5 to 9 on two-dimensional silver stained gels. Major changes in polypeptide patterns of the presence/absence type due to inoculation or genotype were not detected up to 10 days after inoculation (4 days after inoculation shown). Intensity differences were too subtle between nts and Bragg to allow quantitation of these changes. The amount of soluble protein in both Bragg and nts382 decreased 35% from 21.9 (± 1.4) to 14.3 (± 0.7) mg per gram of leaf tissue between one and ten days after inoculation. This decrease may be due to leaf aging and has been observed in other plants e.g., in barley in which leaf proteins decreased 40% due to aging [Huffaker 1982]. There was no significant difference between protein content of inoculated vs uninoculated leaves of Bragg and nts382 (data not shown).

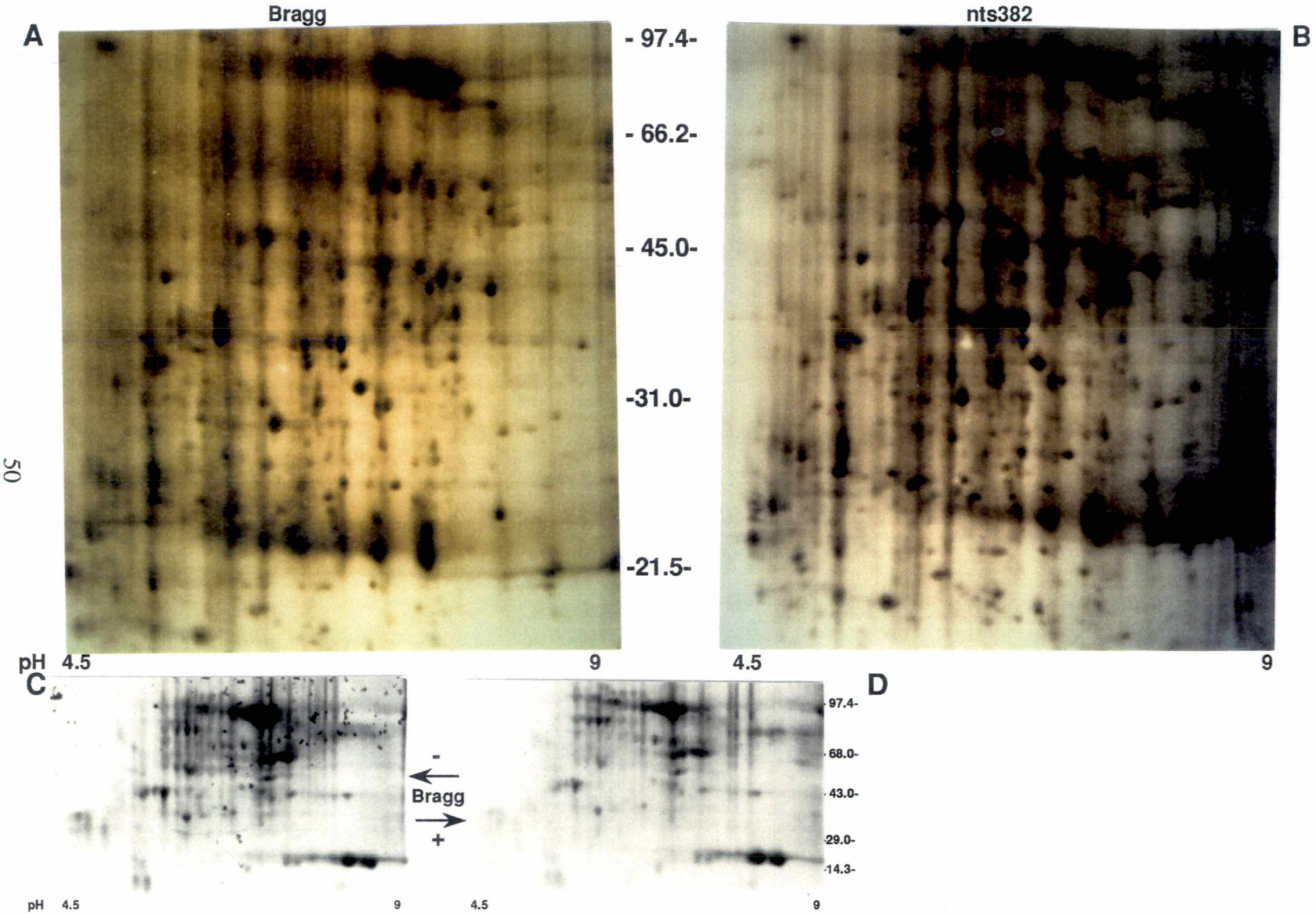
Approximately 55 *in vivo* labeled polypeptides were detected in inoculated (panel D) and uninoculated Bragg (panel C), 10 days after inoculation. No inoculation-related differences were observed in Bragg. Further *in vivo* labeling experiments with both $^{35}SO_4^{-2}$ and ^{14}C-leucine showed no differences between Bragg and nts382 leaf types and inoculated and uninoculated conditions (J. Deckert, unpublished data).

In vitro translations performed on mRNA samples taken from the time of inoculation up to 25 days after inoculation showed no detectable differences due to inoculation [Sayavedra-Soto et al., 1995]. However, two polypeptides associated with changes in plant development tended to decrease by 8 days, then showed a transient peak at around 10 days after inoculation and then decreased gradually. A comparison of the *in vitro* translation patterns twenty-five days after inoculation showed that patterns did not change in the number of polypeptides, but in the level of intensity of their expression. Changes associated with inoculation status were most pronounced at 25 days. Five polypeptides showed changes associated with inoculation between 10 and 25 days after inoculation (DAI) [Sayavedra-Soto et al., 1995].

IV: DISCUSSION

No major changes were detected in total, *in vivo* labeled, or *in vitro* translated proteins before 8 days which is the time point by which autoregulation is fully functional. Changes in proteins detected by *in vitro* translation of mRNA may be correlated to nodulation and subsequent nitrogen fixation. Nitrogen-fixing soybean plants produce large amounts of ureides allantoin and allantoic acid, which after translocation from the root to the leaf are

Figure 3: Comparison of shoot proteins from *G. max* cv. Bragg and *G. max* line nts382 by silver-stained two dimensional gel electrophoresis. Total shoot proteins from Bragg (A) and nts382 (B), were extracted 4 days after inoculation and separated according to Materials and Methods. ^{35}S *in vivo* labeled proteins from inoculated Bragg (D) and uninoculated (C) were extracted 10 days after inoculation and separated according to Materials and Methods.. The pH gradient in the first dimension was from 4.5 to 9. Molecular sizes are indicated.


Bragg
nts382
A
B
C
D
- 97.4-
- 66.2-
- 45.0-
-31.0-
-21.5-
pH 4.5
9
4.5
9
-
Bragg
+
97.4-
68.0-
43.0-
29.0-
14.3-
pH 4.5
9
4.5
9

metabolized or stored. Thus concomitant polypeptide changes are anticipated in such leaf material. Only the analysis of several replicate translations from different plant sets allowed clear discrimination between nodulation/developmental changes and more random environmental changes. This experimental lability may have been the cause of previous reports of protein differences in mutant soybean plants [Gresshoff et al., 1988a,b].

The lack of detectable protein differences between Bragg and nts382 underscores the complexity of the autoregulation response and offers several interpretations.

(a) The infection signal 'Q' from roots may be modified by a constitutive system in the leaves therefore no changes in shoot proteins were seen. A constitutively expressed 66 kD peptide has been found to be associated with pea *sym 5* mutants [Fearn and LaRue 1990]. The mutants show temperature-dependent nodulation, nodulate normally at 12°C but form few or no nodules at normal temperature. The 66 kD protein has an altered pI in the mutant compared to wild-type pea and is expressed in lateral root, tap root and shoot. Its expression was found to be independent of rhizobial inoculation, root temperature or light. Reciprocal grafting experiments showed the mutant phenotype to be associated with the genotype of the roots and not the shoot [Fearn and LaRue 1990]. Nonnodulating mutants that are blocked at the nodule-primordium stage can be restored to normal development if treated with AVG (aminoethoxyvinyl glycine) or Ag^{2+} which inhibits ethylene formation and action, ethylene being a potent inhibitor of nodulation [Guinel and LaRue 1991]. It is possible that the 66 kD protein is involved in the ethylene biosynthetic pathway although no evidence exists for this hypothesis.

(b) Since trifoliolates, primary leaves, petioles and apices were included in the sample used for RNA isolation, the tissue responsible for the regulation of nodulation could comprise a minor proportion of the total shoot sample used. Therefore, quantitatively different mRNA populations or their products may be underrepresented in the analysis. This explanation is supported by some estimates of the number of genes present in higher plants. DNA-RNA hybridization experiments in tobacco suggest that at least 100,000 genes are expressed in the entire plant. About 8000 mRNAs are common to different cell types, but each organ (leaf, root, flower) contains a specific set of mRNAs found exclusively in the polysomes of cells from that organ [Goldberg 1986]. Considering that 35% of ovary nuclear transcripts are not found in other nuclei and that 45% of tobacco single copy DNA is transcribed, of which only 30% is transported suggests that there may be about 16,000 genes found in any

cell type [Goldberg 1986]. If approximately half to a third of these mRNAs are expressed as protein, one can estimate that the number of mRNAs being expressed about 5000 to 8000. *In vitro* translation in this study detected approximately 150 proteins on mini gels which could be increased by using longer gels and larger sample volumes. In another study of *in vitro* translated products from *G. max* cvs Verdon and Arrow shoots, two-dimensional gels detected 566 radiolabeled translated products [Cabané et al., 1992]. This suggests that even with sensitive detection techniques, 10% (or less) of the expressed mRNA products can be detected. The protein changes involving the nts shoot could be associated with low levels of mRNA and /or rapid turnover of mRNA. Therefore unless the protein change involved a mRNA that is present in large numbers, and/or slow to turnover, the inclusion of various tissues could mask the detection of low concentrations of mRNA.

(c) The polypeptides involved in shoot-controlled autoregulation are expressed at a level that is below the detection limits of gel electrophoresis. The lack of differences in total protein suggests that the effect of the *nts* mutation is not absolute and that the protein differences between Bragg and nts shoots may be too subtle to detect.

(d) Since no *de novo* protein changes were found by *in vivo* labeling, this may indicate either a short half-life of the protein or that the altered product in the supernodulation mutant (i.e. the *nts* gene product) is post-translationally regulated. Most of the sulfate arriving in the leaves is reduced and the reduced sulfate is translocated to various sinks such as rapidly turning over cysteine, slowly turning over cysteine, protein, precursors and methionine [Giovanelli et al., 1980]. Although nearly all of the intracellular cysteine and methionine is in the form of protein, flux measurements in *Lemna* cells indicate that the relative fluxes between the two cysteine pools are three times higher compared to all the other pools [Giovanelli et al., 1980]. In case of *Lemna* cells, compartmental analysis indicates that the pool of free sulfate in the cytoplasm is only 1%, but free sulfate in the vacuole is 25% of the total sulfur of the cells [Cram 1980]. If the half life of the protein is less than the turnover $^{35}SO_4$, such a protein would not be detected.

(e) The *nts* mutation as well as the onset of autoregulation are not associated with major pleiotropic changes in the leaf biochemistry and physiology. Alternatively, there may be no gene expression changes in soybean leaves upon inoculation.

I favor interpretation (e) since the product of *nts* might be a regulatory protein which are normally present at very low concentrations and therefore would be very difficult to detect.

In hindsight, when no differences were found with a variety of techniques, experiments looking specifically at differences in phosphorylated states of proteins between Bragg and nts could have been explored further. If the auxin burst theory is correct [Gresshoff 1993] then more emphasis should be placed on looking for protein differences in the root rather than the shoot as a target for auxin-related effects and increasing the chances of finding a target of the *nts* gene.

The data suggest that the isolation of gene sequences involved in autoregulation and supernodulation through detection of polypeptide differences is difficult and that other approaches, for example, positional cloning using a closely linked molecular marker, large fragment isolation and functional complementation must be explored.

PART III

DETECTION OF DNA POLYMORPHISMS BY DNA AMPLIFICATION FINGERPRINTING

The three major DNA profiling protocols that use arbitrary primers are DAF [Caetano-Anollés et al., 1991b], RAPD [Williams et al., 1990], and AP-PCR [Welsh and McClelland, 1990]. The three methods differ in terms of their reaction conditions, separation and the detection system used. Important distinctions between the three techniques are discussed below, a more detailed review can be found in Caetano-Anollés et al., [1991a]. RAPD uses primers that are 9 or 10 nucleotides long, the amplification products are separated on agarose gels and visualized by ethidium bromide staining. AP-PCR differs from RAPD and DAF mainly in two ways. First, primers between 10 and 20 nucleotides in length are used. Second, the stringency of cycling parameters is varied during the amplification. Primary amplification products are obtained under the first two cycles of low stringency followed by high stringency cycles. Third, amplified products are separated on polyacrylamide gels and detected by autoradiography of radiolabeled products. DAF differs from these methods in three major respects. The amplification reaction uses very short primers (as short as 5 nucleotides can be used [Caetano-Anollés et al., 1991a]), and the amplified products are detected on a polyester backed silver-stained polyacrylamide gel containing 7 M urea [Bassam et al., 1991]. More importantly, unlike RAPD or AP-PCR, DAF uses a high primer/template ratio of concentration in the amplification reaction. These factors contribute to the ability of DAF to detect a large number of amplification products, the ability to tailor profiles of varying complexity, and successful amplification above an annealing temperature of 45°C.

The detection methods used in the three methods differ with respect to resolution, ease of performing and speed. Although sequencing gels used in AP-PCR are highly sensitive, running radioactive sequencing gels is problematic and slow. The commonly used ethidium bromide stained agarose gels are easy and fast but are limited by decreased sensitivity and low resolution of smears. Moreover, they are not archivable. Silver-stained DAF gels offer a compromise in that they are fast and easy to run. They also work as a long term depository of DNA profiles and a source of amplified fragments as DNA can be recovered from dried gels [see part VI and Weaver et al., 1994].

DAF gels differ from sequencing gels in that they are shorter (10 cm long), the samples are not denatured prior to electrophoresis and are run at room temperature. Staining is based on a photochemically derived staining procedure, that is different from the previously reported

alkaline and acidic methods of silver staining [Merril, 1990]. It is characterized by low background staining without loss of contrast, has fewer staining steps, and is faster than previously published protocols [Bassam and Caetano-Anollés, 1993]. Fingerprints from a variety of organisms like bacteria, bermudagrass, monkey, human and soybean showed a high quality resolution using this technique [Caetano-Anollés et al., 1991b].

One major problem encountered with RAPD reactions has been reproducibility of fingerprinting patterns. This has raised concerns regarding the use of RAPDs for genotyping purposes. It is important to distinguish between non-reproducible bands resulting from an altered genetic background (e.g., F2 plant, see part IV) and those resulting from errors in the amplification process itself due to reaction conditions. In a study with 23 celery cultivars [Yang and Quiros, 1993] only 30% of the polymorphic bands were reproducible. In *Arabidopsis* a high density map containing 252 RAPD markers was constructed using a recombinant inbred mapping population [Reiter et al., 1992]. Out of 392 polymorphic markers, 176 markers were not reproducible. In these studies, the source of these non-reproducible bands was not clear. Non-optimal reaction parameters contributing to artifactual variation in RAPD technology have been shown to be primer to template ratio, annealing temperature, magnesium concentration [Ellsworth et al., 1993], different brands of thermocyclers and different brands of *Taq* polymerase [MacPherson et al., 1993, Meumier and Grimont 1993]. Before arbitrary primer technology (APT) markers can be used for fingerprinting and mapping purposes, it is important to demonstrate their reproducibility by establishing an optimum set of reaction conditions.

The purpose of this chapter is (1) to establish DAF as a *bona fide* method for detecting polymorphisms in terms of reproducibility and increased resolution over existing methods of detection and (2) to establish a standard curve for silver staining of DNA fragments on polyacrylamide gels since no published data exist showing the linearity and limits of detection of this newly developed stain.

II: MATERIALS AND METHODS

The DAF reaction requires template, nucleotides, buffer, Mg ions, enzyme, and primer. Since all these factors can influence the reaction it was important to optimize these parameters. The first step is to isolate the template DNA from a suitable plant tissue. All DAf reactions described in this dissertation use young trifoliolates as the starting material.

The DNA is then amplified, separated on polyacrylamide gels, and stained with silver. Details of the procedures for all these methods is given in part VIII. Optimum reagent concentrations are obtained by using a range of concentrations for amplification in which reproducible profiles can be obtained. For determining sensitivity and linearity of the silver staining procedure, molecular weight standards are used since their concentration and size is known. Laser densitometric scanning is a way to quantify the intensity of the bands seen on a gel by calculating the height or the area of the peaks.

III: RESULTS

Optimum template and enzyme concentration

Enzyme and template concentrations were optimized in the DAF reaction with the objective of obtaining reproducible profiles for a range of template and enzyme concentrations. In order to obtain a DAF profile, the template DNA was first extracted, amplified, and the products were then separated on polyacrylamide gels that were silver stained. Details of protocols for all three are presented in the Materials and Methods section. Reproducible patterns were generally obtained between 0.04 - 0.06 units of *Taq*® polymerase per µl reaction volume (Fig. 1, lanes 9 and 10). Increase in enzyme concentration to 0.08 units per µl reaction volume produced a smear (lane 11) and a decrease to 0.02 units of enzyme per µl of reaction volume produced no products (data not shown).

The optimum template concentration for obtaining highly reproducible fingerprints for soybean DNA was found to be between 0.2 and 0.4 ng per µl reaction volume (lanes 3-5). The effect of template concentrations on reproducibility of profiles outside this range is shown in lanes 1, 2, 6, 7, and 8. Decreasing the DNA concentration to 0.05 ng per µl gave no bands (lane 2). Increasing the concentration to 2 ng/µl (lane 6), 20 ng/µl (lane 7) and 140 ng/µl (lane 8) gave altered profiles in terms of different sized bands or complete loss of bands. At 2 ng/µl some common bands are present but additional bands of different sizes are produced. At 20 ng/µl, bands completely different in size from those at 0.2 or 0.4 ng/µl are produced with primer 8.6a. Other primers tested produced no amplification products at 20 ng/µl (data not shown). At 40 ng/µl (lane 8), low molecular weight products disappeared only high molecular bands were amplified. Template concentrations between 0.2 ng/µl to 0.4 ng/µl produced consistent profiles for all the arbitrary primers tested (see part IV). Other reaction parameters like primer, nucleotide, Mg^{2+} concentrations and cycling parameters have been optimized in other studies with soybean [Caetano-Anollés et

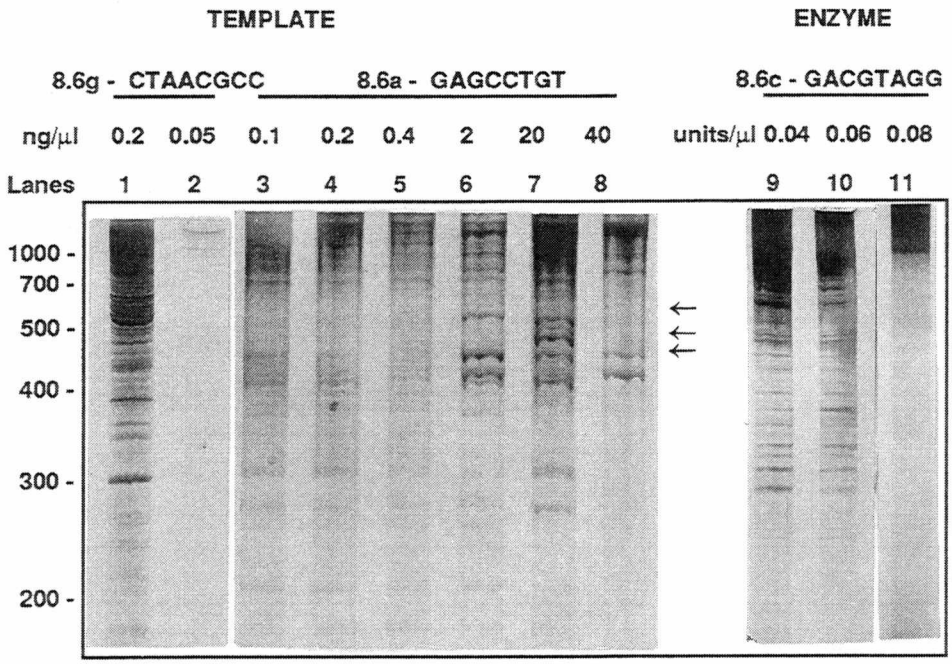


Figure 1: Optimization of template DNA and enzyme concentration for amplification. Concentration of *G. soja* DNA was varied from 0.05 to 40 ng/µl reaction volume (see materials and methods). Amplification profiles were consistent at 0.1, 0.2 and 0.4 ng/µl (lanes 3-5). At 2 ng/µl (lane 6), 20 ng/µl (lane 7), and 40 ng/µl (lane 8) extra bands (arrows) were produced. At 0.05 ng/µl (lane 2) no bands were observed (compare with lane 1 at 0.2 ng/µl). Consistent profiles were obtained for *G. max* with enzyme concentrations at 0.04 units (lane 9) and 0.06 units (lane 10) per µl reaction volume. Amplification using 0.08 units /µl resulted in a smear (lane 11).

al., 1991] and turfgrass [Weaver 1993], and these were used without further optimization. Weaver [1993] showed that for turfgrass, 1.5-3 mM Mg^{2+}, 2.4-7.2 μM primer and 40-320 μM nucleotides was optimum for the DAF reaction.

Reproducibility of DAF

Figure 2 demonstrates consistency of DAF profiles between two different extractions and three individual amplifications on three separate gels of *G. soja* and *G. max*. DNA for amplifications was extracted six months prior to use and the amplifications were performed nine months apart. Monomorphic and polymorphic bands were reproducible in all three sets of DAF profiles. Between 100 and 1000 bp, *G. soja* (lanes S) and *G. max* (lanes M) profiles showed sixteen monomorphic bands (9, strong and medium intensity; 7, minor) and three polymorphisms at 1000 bp (class I), 250 bp (class II), and 200 bp (class II). Since polymorphisms of different intensities are routinely detected on DAF gels (see part IV), it was useful to adopt a scheme for classifying polymorphic bands based on their intensity. The classification was not based on quantitation using a laser densitometer for instance, but was done on a visual basis. Class I polymorphisms are scored as bands of strong intensity and approximately equal in intensity to the major monomorphic bands in the DAF profile. Class II polymorphisms are bands of medium intensity compared to poly- or monomorphic bands of strong intensity. Class III are bands of least intensity in the profile (see part IV).

Comparison of detectability of polymorphic band detection on silver-stained DAF gels and ethidium bromide-stained agarose gels

To compare the sensitivity of detection of polymorphisms on silver-stained polyacrylamide gels, amplification products of *G. soja* and *G. max* with primers 8.6n, and 8.10b were also separated on agarose gels and stained with ethidium bromide (Fig. 3). With 8.6n, *G. soja* (lane 1) and *G. max* (lane 2) showed almost identical DAF profiles composed of six monomorphic bands ranging from 0.19 to 1.9 kb and one faintly detectable polymorphic marker at 240 bp when resolved on a 2% agarose gel. In comparison, between 200 and 1000 bp, on a silver-stained polyacrylamide gel, *G. soja* (lane 3) and *G. max* (lane 4) showed polymorphisms at 240 bp (class I) and 205 bp (class III) and 19 monomorphic bands (11, strong and medium intensity; 8, minor). Twenty five μl of amplified products were required to detect the poorly resolved faint band at 240 bp on the agarose gel, whereas 3 μl of amplified products was sufficient to detect two markers at 240 and 205 bp on

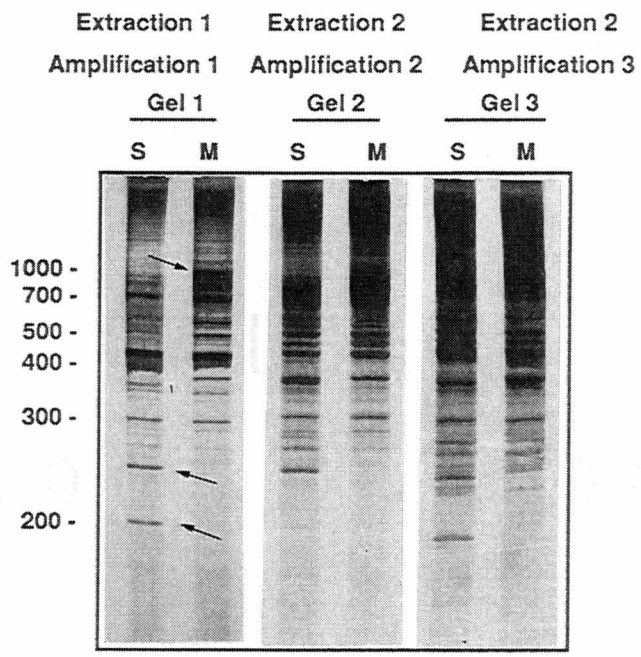


Figure 2: Reproducibility of DAF profiles.
Consistent DAF profiles were obtained with primer 8.10b for *G. soja* (S) and *G. max* line nts382 (M) from two different extractions of template DNA, three individual amplification reactions and three different gels. Monomorphic and polymorphic bands (arrows) were reproducible. See materials and methods for DNA extraction, amplification conditions and gel electrophoresis.

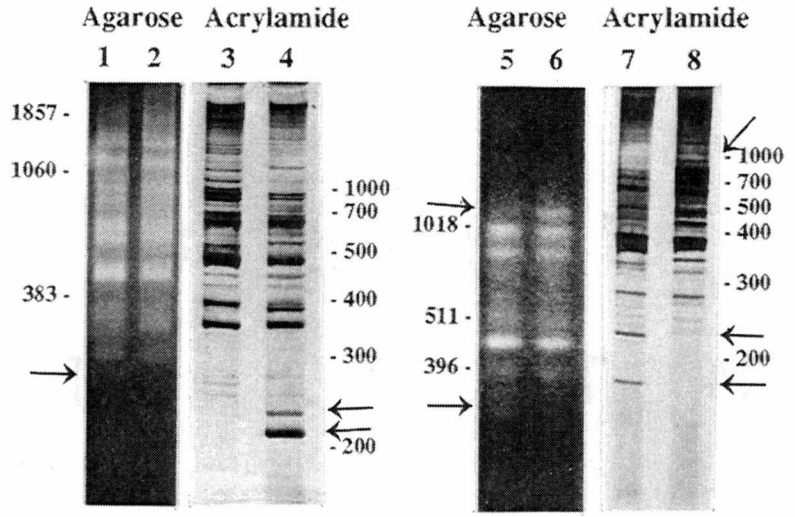


Figure 3: Comparison of DAF profiles on ethidium bromide-stained agrose gels and silver stained-polyacrylamide gels.
G. soja (lanes 1, 3, 5, 7) and *G. max* line nts382 (lanes 2, 4, 6, 8) were amplified using primer 8.6n - GTAACGCG (lanes 1-4) and 8.10 b - GCCCGCCC (lanes 5-8).
Polymorphic bands are indicated by arrows. See text for further explanation.

polyacrylamide gels. The 205 bp marker failed to be detected on agarose. In case of primer 8.10b, *G. soja* (lane 5) and *G. max* (lane 6) showed six monomorphic bands (of these three are minor) ranging from 0.22 to 1.2 kb. Two markers at 1200 bp (intense) and 280 bp (very faint) were detected on a 1.8% agarose gel. In comparison, between 100 and 1000 bp, *G. soja* (lane 7) and *G. max* (lane 8) showed sixteen monomorphic bands (9, strong and medium intensity; 7, minor) and three markers at 1000 bp (class I), 250 bp (class II), and 200 bp (class II) on polyacrylamide gels. The difference in molecular size of the polymorphisms detected is due to differences in mobility of bands on agarose and polyacrylamide gels. To summarize, class I polymorphisms on DAF gels were detected as strong or faint bands on ethidium bromide-agarose gels, whereas most class II and class III bands on polyacrylamide were undetectable on agarose gels. Compared to agarose, DAF gels detect a two to three fold increase in polymorphic bands compared to agarose gels.

Limits and linearity of DNA fragment detection on silver-stained DAF gels

Figure 4A shows molecular size markers from 200 to 1000 bp separated on a silver-stained 5% (w/v) polyacrylamide gel. The amount of DNA in each band ranged from 0.05 ng/band to 30 ng/band and all markers were equal in GC content [BioVentures manual]. A laser densitometric scan (Fig. 4B) of the gel shows that peaks corresponding to the fragments are separate and distinct up to 1.5 ng/band. Further decrease in the amount of DNA up to 150 pg/band per band produced scans with a high signal to noise ratio. The scans in Fig. 4B had to be reduced in size for presentation purposes, therefore peaks representing the fragments are not clearly distinguishable at 0.5 and 0.15 ng/band. However, the gel analysis software was able to assign peaks and absorbance values. Consistent with no visible bands, at 50 pg/band (scan not shown) the densitometer peaks were indistinguishable from baseline readings.Visual inspection was as good or better than laser densitometry (compare Fig. 4A with 4B). The peaks at 1.5 ng/band are barely visisble in the scans shown in fig. 4B whereas they are clearly visible in fig. 4A

Quantitative analysis of the absorbance data from the densitometric scan in Fig. 4B is shown in Fig. 5. A graph of the intensity of molecular size bands versus the amount of DNA in each band (Fig. 5A) shows that the intensity of bands from 0.5 to 30 ng/band for all molecular sizes (only 200 and 700 bp shown) is curvilinear. However, as shown in Fig. 5A (inset), the intensity of the DNA fragments is linear from 0.5 to 5 ng/band. The intensity of bands at 150 pg/band is outside this linear range for all sized bands (data not

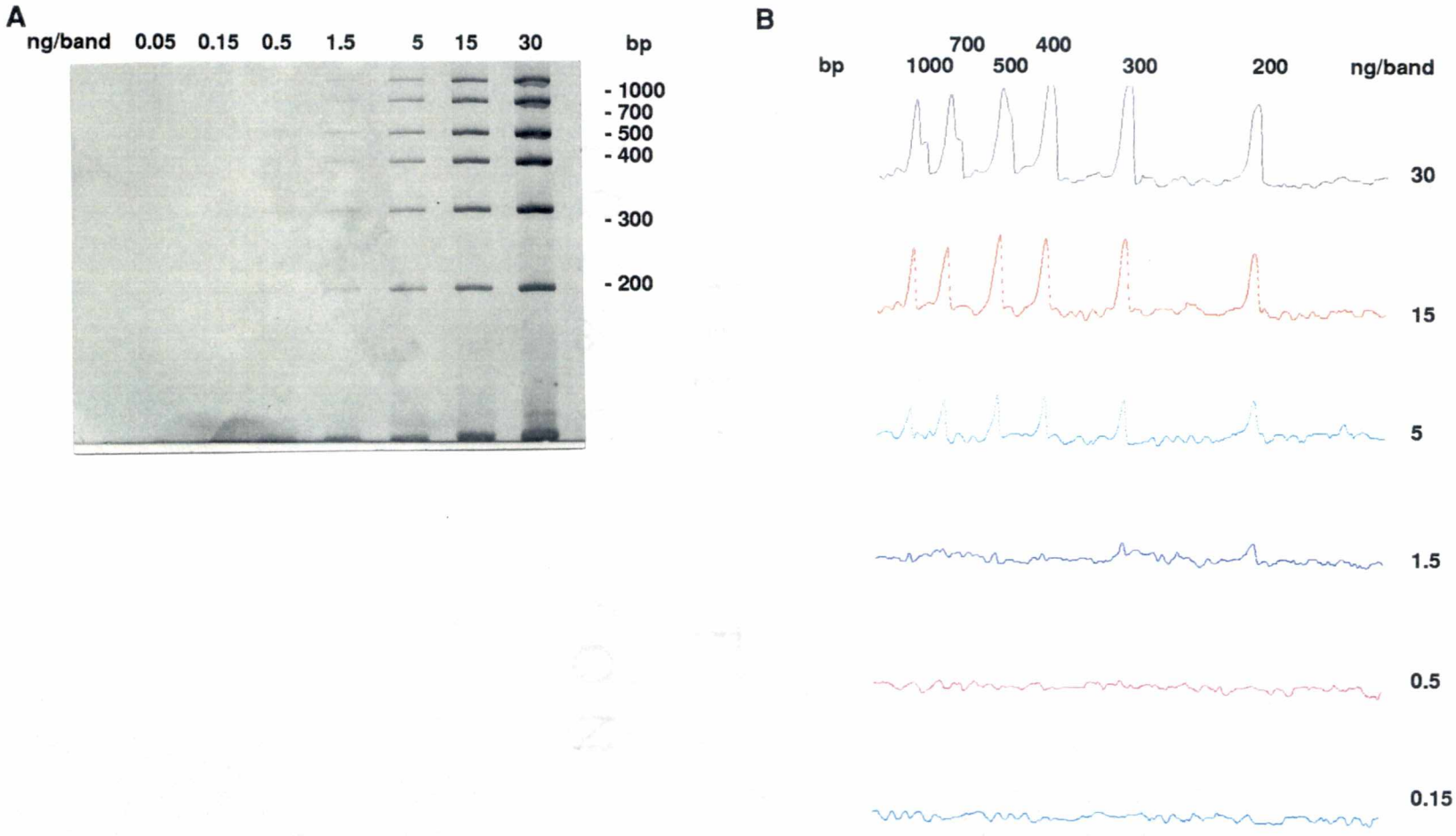


Figure 4: Limits and linearity of DAF silver staining.
Panel A: Molecular size standards between 200 and 1000 bp varying from 0.05 ng/band to 30 ng/band were separated on a silver stained DAF gel.
Panel B: Laser densitometric scan of the gel shown in lanel A (0.05 ng/ band scan not shown)

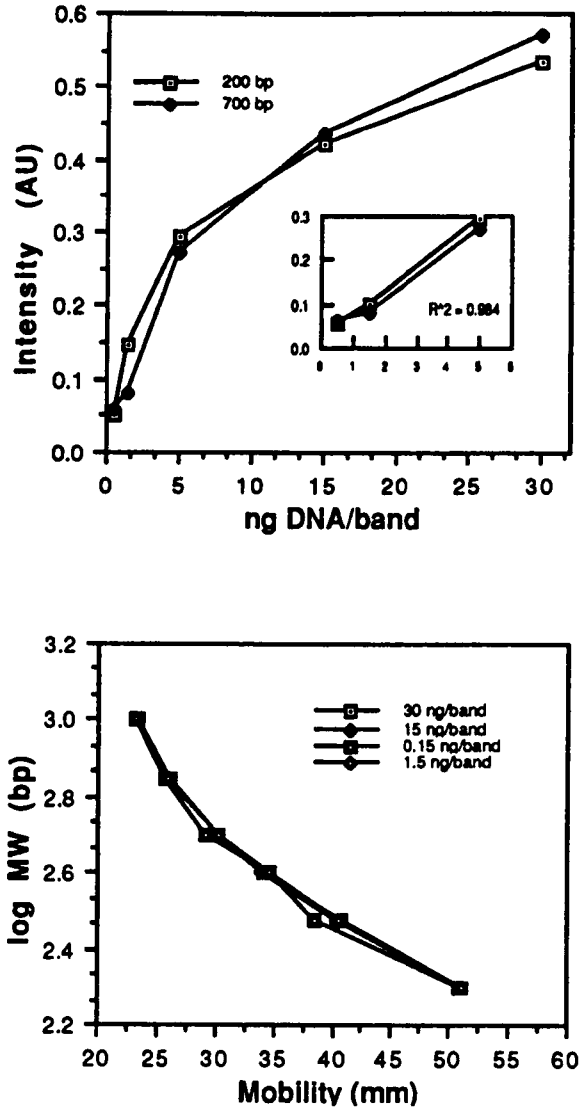


Figure 5: Intensity and mobility as a function of size and mass of silver stained DNA fragments. Data from the densitometric scan in Fig. 4B was used to determine the standard curve (Panel A) of intensity vs amount of DNA ranging from 0.15 to 30 ng/band and to determine the mobility of DNA fragments as a function of molecular size (Panel B).

shown). The average standard error for the intensity of all size bands measured from a replicate standard curve was ± 3%. The replicate experiment (data not shown) also showed that although the absolute intensity of the bands varied between gels depending on the time of development, the intensity differences within the same gel remained constant.

A graph of the mobility of DNA fragments versus the logarithm of their molecular size shows that different amounts of DNA up to 30 ng/band exhibit identical mobility (Fig. 5B). The average standard error for mobility of all size bands was less than ±1%. The effect of GC content of a fragment on its mobility was analyzed by separating *Alu*I restricted pBR322 fragments on a 5% polyacrylamide gel. The different fragments ranging from 40-60% GC showed less than 1% variation in mobility (data not shown).

Quantitation of bands in DAF profiles

Figure 6 shows a densitometric scan of amplification profiles of DNA from *G. soja* and *G. max* using primer 8.6d (GT AAC GCC). This gel was chosen as a representative profile for amplification products of varying intensity in DAF profiles. Visual inspection of the gel in Fig. 6 showed ten bands in *G. soja* (4 strong, 2 medium and four very faint intensity) and eleven bands in *G. max* (one extra faint band). Band sizes were 320, 285, 250, 200, and 180 bp for peaks 1 through 5 respectively. The densitometric scan showed 5 peaks (arrows) for *G. soja* and 6 peaks for *G. max* that varied in intensity. Fragment number 2 was detected as a split peak in *G. max*. This lack of sensitivity of the densitometer was assumed to be a limitation of the gel scanning software. The highest intensity peak was peak 1 in *G. soja* (0.518 AU) and the faintest peak was peak 5 in *G. soja* (0.057 AU). Typical amplification products, therefore, fall within the values measured for the standard curve, 0.6 AU at 30 ng/band and 0.05 AU at 0.5 ng/band (Fig. 5). Peak 5 is an example of a band that shows a five-fold difference in intensity between *G. soja* and *G. max* but is not considered polymorphic.

IV: DISCUSSION

Results shown in this study indicate that the DAF reaction is highly reproducible between DNA extractions, amplifications, and gels. Amplifications performed nine months apart showed identical DAF profiles, polymorphic bands of both strong and weak intensity were consistently produced. Moreover, extractions for these amplifications were performed six months prior to use indicating that the extracted genomic DNA is stable for at least this

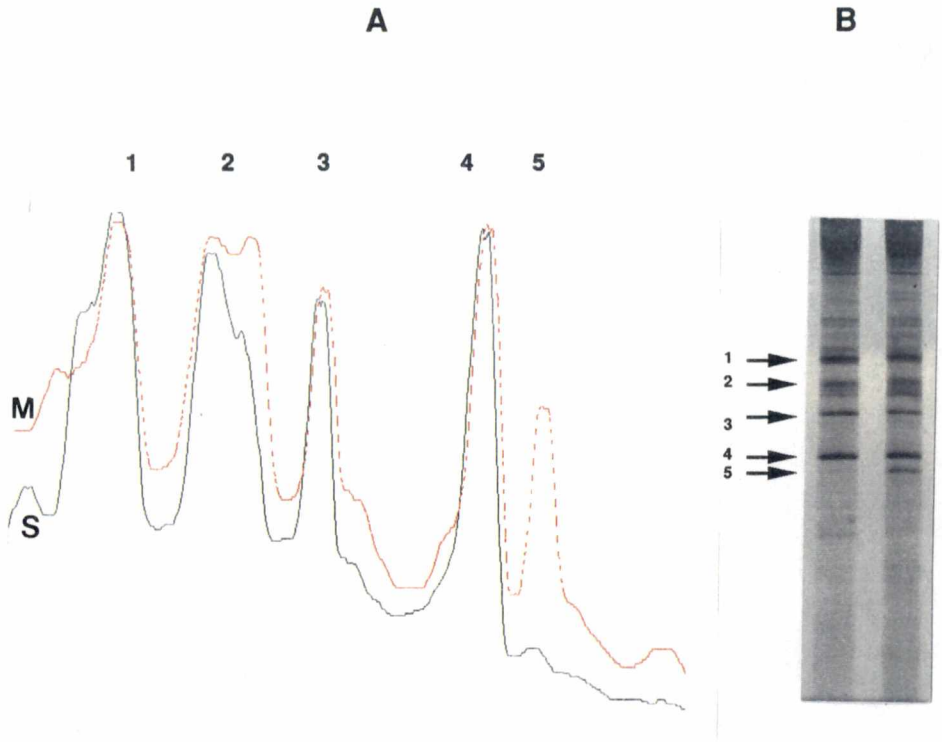


Figure 6: Quantitative analysis of DAF products.
Amplification profiles (Panel B) from *G. soja* (S) and *G. max* (M) using primer 8.6d- GTA ACG CC were scanned by laser densitometry (Panel A). Dashed line in panel A corresponds to *G. max* profile in panel B and numbered peaks correspond to numbered bands.

length of time. Evidence for reproducibility also comes from a study by Gresshoff and MacKenzie [1994] who showed that patterns from ten individual *Glycine max* cv. Bragg plants produced identical fingerprints consisting of 35 bands. The same study showed that the DAF reaction when performed by two different people using their own reaction materials produced identical profiles, further attesting to the reproducibility of DAF. The reliability of DAF is significant in terms of transfer of technology to other laboratories and lab personnel.

High information containing DNA profiles of *G. soja* and *G. max* showed monomorphic and polymorphic bands of different intensities. Depending on the intensity, DAF markers were classified into three classes. As yet there is no explanation for the different intensity of bands. However, primer and template engineering experiments have suggested that secondary structure within the amplicon, or flanking it, can affect the efficiency of amplification and give rise to bands of different intensity [Caetano-Anollés et al., 1992a]. Compared to ethidium bromide stained agarose gels, silver-stained polyacrylamide gels detected a higher number of DAF products and three times as many DAF markers.

The optimum window of enzyme and template concentration required for consistent DAF profiles with discrete, scorable bands were established in this study. Altered profiles were produced outside the optimal window of template concentrations from 0.1 to 0.4 ng per µl reaction volume. Some bands were less sensitive to template concentration and amplified at all concentrations tested whereas other bands were more sensitive and varied in consistency. Low ratio of primer to template conditions produce longer products, an observation that has been made in several cases [Weaver 1993 and Weeden 1992]. In other words some amplification products can be amplified under different ratios (low to high) of primer to template, whereas other amplification products are more sensitive to this ratio. Explanations for the observed sensitivity of some bands could include competition, primer-template mismatch, and the number and type of sequence the band originates from in the genome. Also the presence of excess primer could eliminate selective amplification of products due to limiting primer concentrations. The optimum range of enzyme concentration may vary depending on the batch of enzyme used. Testing every batch of enzyme for an optimum concentration using a standard set of conditions is therefore recommended if lack of consistency during amplification is found. This was not found to be necessary for most amplifications (see part IV).

For a primer that is 8 nucleotides long, the concentration of primer in the DAF reaction is 3 μM. At 10 ng of template this corresponds to a primer to template mass ratio of 19. For standard RAPD reactions with 1 ng/μl reaction volume of template DNA and 0.2 μM primer [Williams et al., 1991], the primer to template mass ratio is 0.52. There is a near 40 fold difference in the primer to template ratio between DAF and RAPD reaction conditions. The importance of primer to template ratios for reproducible profiles is further confirmed by the fact that reproducibility of RAPD reactions was improved by using template concentrations in the range of 3 to 30 ng/25 μl and primer in the range of 0.2 to 1 μM, concentrations that are closer to DAF conditions [Weeden et al., 1992; Smith and Chin 1992].

The control of the amount and purity of DNA is critical for obtaining reproducible results. The optimum amount of DNA for different genomes is not identical but may vary, e.g., 3 - 4 ng DNA per μl of reaction volume is optimal for rice, 0.14 - 0.8 per μl for banana [Munthali et al., 1992] and 1 ng/μl is optimal for bacteria [Bassam et al., 1992].

The results described in this section and those from optimization studies by Weaver [1993] suggest that some components of the DAF reaction like template, primer and enzyme concentrations are more critical and have a fairly narrow window of optimal conditions whereas other components like nucleotide and Mg^{2+} concentration have a larger window. However, these ranges of concentrations are not so narrow so as to seriously impair reproducibility of DAF patterns by researchers in other laboratories.

Caetano-Anollés et al., [1991b] have shown that it is possible to obtain and detect up to 100 DAF products on a silver-stained gel. Figures 2 and 3 (also see part IV) show approximately 25 bands between 100 and 1000 bp on a 5% acrylamide gel. The number of products detected can be increased either by running longer gels or by using multiple mini-gels of different acrylamide concentrations. Silver stained sequencing gels detected on average 3-5 times higher number of DAF products compared to mini-gels. For example between 300 and 500 bp, sequencing gels detected 35 and 46 fragments while mini-gels detected 7 and 8 fragments with two different primers. (Shane Abbitt, pers. comm.). A comparison of silver-stained polyacrylamide gels and ethidium bromide stained agarose gels showed that polymorphic bands comprised 6.6 % of the total number of bands on agarose gels (considering only the strong polymorphism), whereas 10% of the total bands were polymorphic on acrylamide gels (considering class I and II bands only). The number of polymorphisms detected on silver-stained DAF gels was three times higher compared to

ethidium bromide stained agarose gels.

Bassam et al., [1991] showed that by silver staining amplification profiles, DNA concentrations up to 1 pg/mm^2 (length x thickness) band area could be visualized, an improvement of up to 30 fold over previously reported methods of silver staining [references can be found in Bassam et al., 1991]. Dependence of fragment size on band intensity has been observed for silver staining with lower sized bands staining more intensely than higher sized bands [Bassam et al., 1991 and Beidler et al., 1982].

In this study, the DAF silver staining method showed a limit of detection close to 150 pg/band or 75 pg/mm^2 but showed no dependence of size on staining intensity. Although the exact cut-off limit for band detection was not determined in this study, longer development of the gel may lower this limit slightly. These detection limits are applicable under the conditions of this particular gel (Fig. 4A) , however, this low background is the best one for optimum resolution of DAF products. The DAF silver staining method is linear from 150 pg to 5 ng/band similar to the acidic method of silver staining [Beidler et al., 1982] which is linear between 0.25-4 ng/band for fragments between 300-1100 bp.

DAF gels do not show abnormal mobility for (1) fragments of any particular size, (2) fragments containing variable amounts of GC (40 - 60%), and (3) variable amounts of DNA up to 30 ng/band. Quantitative estimation of soybean DAF products by Gresshoff and MacKenzie [1994] showed that band mobility on average varied by 2% which is consistent to the 1% variation observed in this study. The Gresshoff and MacKenzie study also showed that although the intensity of amplified products varied by an average of 51% between different amplifications, intensity differences within the same gel remained consistent. Polymorphic bands were extremely reproducible between DNA extractions, amplifications and gels similar to the results presented in Fig. 2.

In conclusion, DAF is reproducible, reliable, highly sensitive, shows no dependence of silver staining intensity on size or GC content of the DNA fragment, and is suitable for screening parental genotypes in order to detect the maximum number of polymorphisms.

PART IV*

INHERITANCE OF POLYMORPHIC FRAGMENTS GENERATED BY DNA AMPLIFICATION FINGERPRINTING AND THEIR USE AS GENETIC MARKERS IN SOYBEAN

* Prabhu RR, Gresshoff, PM: Inheritance of polymorphic markers generated by DNA amplification fingerprinting and their use as genetic markers in soybean. Pl Mol Biol 26: 105-116 (1994)

I: INTRODUCTION

Soybean has been shown to exhibit a low level of genetic diversity. Only 40% of low copy random *PstI* genomic clones detected polymorphisms between *G. soja* PI 468.916 and *G. max* line A81-356022. Approximately 20% of the clones detected repeated sequences [Keim and Shoemaker 1988]. Pedigree analysis of North-American soybean cultivars showed that 70 to 80% of their genomes were derived from only 17 accessions [Shoemaker et al., 1992]. In a study of 58 soybean accessions [Keim et al., 1989] a higher degree of genetic diversity was detected between *G. soja* and *G. max* (33%), as compared to within cultivated varieties of soybean (16%). A random sample of 32 RFLP probes detected polymorphisms with a frequency from 0 to 72% between pairs of genotypes from a set of 51 genotypes representing U.S. germplasm [Shoemaker et al., 1992]. This increases the difficulty of using RFLP markers even though RFLP markers are almost unlimited in number. Alternative molecular markers like microsatellites have been found to be frequent and highly polymorphic in eukaryotic DNA. However, the technique requires prior sequence information to 'drive' the PCR reaction. It is clear that molecular markers for genome mapping need to be revealed by multiple technologies.

Different strategies can be used to increase the probability of detecting a polymorphism among individuals using single arbitrary primer methods. First, organisms that are genotypically distant can be selected for the mapping population. Examples of such crosses can be found in the construction of maps of soybean, *G. max* (A81-356022 x *G. soja* PI 468.916 (cultivated x wild type) [Keim et al., 1990b], common bean, *Phaseolus vulgaris* (XR-235-1-1) x *Phaseolus vulgaris* cv. Calima (Mesoamerican breeding line x Andean cultivar) [Vallejos et al., 1992] and *Brassica*, (*Brassica oleracea* var. italica x *Brassica oleracea* var. capitat [Slocum et al., 1990]). Second, an optimized amplification reaction that produces a large number of products is required to ensure that a large number of loci are scanned in the genome for potential differences. Third, a detection system capable of revealing the polymorphisms reliably and accurately is used.

DAF has been shown in part III to be reliable and capable of detecting an increased number of polymorphisms. Used in combination with parents that are genotypically distant, DAF is suitable for detection of polymorphisms. Since APT markers are produced by a different mechanism compared to RFLPs and microsatellites, polymorphisms detected by single primer methods may define previously unmarked regions in the genome. The use of APT

markers eliminates the need for Southern blots, the need for large amounts of restrictable DNA and requires no prior knowledge of the template sequence as primers of arbitrary sequence are used. While DAF patterns have been useful for DNA profiling, little evidence existed as to their utility as genetic markers. DAF generated markers may be considered useful for the generation of high density maps and map-assisted breeding only if inheritance is confirmed. The objective of this study was (1) to evaluate the capability of DAF to reveal heritable molecular markers in soybean using a cross between *G. soja* and *G. max*, and (2) to show that these markers can be linked to existing markers on a soybean map.

II: MATERIALS AND METHODS

Two different mapping populations are used in this part. *G. soja* and *G. max* , F1 and F2 is used to screen different primers to obtain DAF polymorphic markers and to study the inheritance of these markers. Minsoy, Noir 1, and RI lines are used to map DAF polymorphisms. The focus of this part is to show mapping of DAF markers in the soybean genome, not necessarily linked to the *nts* region. DNA isolation, DAF reaction, DAF data analysis, and method of molecular mapping is described in part VIII, Materials and Methods.

III: RESULTS

Detection of polymorphisms between G. soja and G. max

DAF profiles of total DNA from *G. max* and *G. soja*, visualized by silver staining of PAGE-separated amplification products, were obtained with 26 oligonucleotide primers (8-mers) of varying GC content. Fig. 1 shows DAF profiles from *G. soja* and *G. max* with four representative 8-mers of varying GC content from 62.5 to 100%. DAF products ranging from 50 to 1000 bp and ranging from light to dark intensity were resolved within each DAF profile. Depending on the intensity, DAF markers were visually scored into three classes. Class I DAF markers were scored as bands of strong intensity and approximately equal in intensity to the major monomorphic bands in the DAF profile. Class II DAF markers were bands of medium intensity compared to poly- or monomorphic bands of strong intensity. Class III DAF markers were bands of least intensity in the profile. DAF markers belonging to more than one class were detected with most primers. For example, as shown in Fig 1, primer 8.6n amplified class I and III DAF markers with *G. soja* (GS) and *G. max* (GM), primer 8.7h amplified class I DAF markers and primer 8.10b amplified

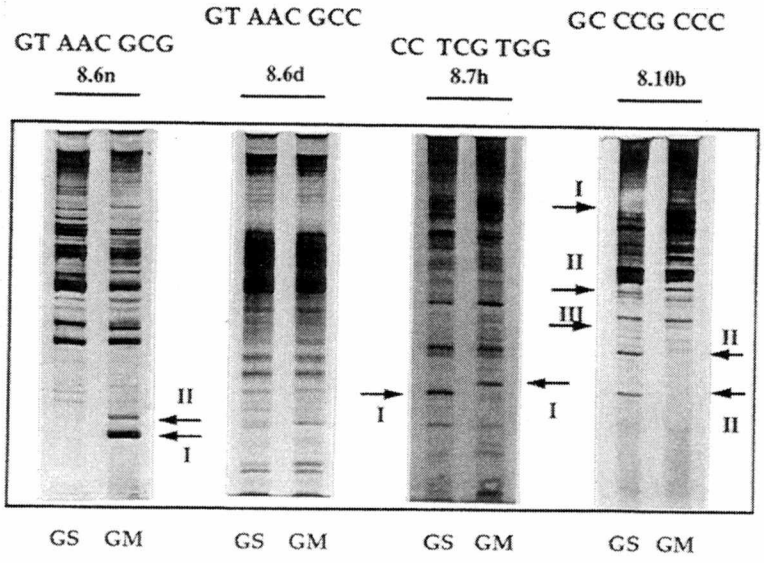


Figure 1: Different classes of polymorphisms between *G. soja* and *G. max*.
Silver stained polyacrylamide gels comparing DAF profiles and AFLPs between *G. soja* (lanes 1, 3, 5, 7) and *G. max* (lanes 2, 4, 6, 8) using primers of varying GC content from 62.5 to 100%. See materials and methods for amplification conditions and gel electrophoresis. The primer and its sequence used for amplification is indicated at the top of the panel. Depending on the intensity, polymorphic bands indicated by arrows, are divided into three classes. I = Class I AFLP of strong intensity, II = Class II AFLP of medium intensity, III = Class III AFLP representing minor amplification products. Note that primers 8.6n (GT AAC GC**G**) and 8.6d (GT AAC GC**C**) are closely related differing only in one nucleotide, yet show differing profiles. GS = *G. soja*, GM = *G. max* line nts382. Profiles for primers 8.6n and 8.10b are used again for comparison in fig. 2 and fig. 7.

class I and class II DAF markers. Due to the highly sensitive nature of the silver stain, verified monomorphic faint bands sometimes appeared polymorphic in the photograph. However, when DAF products were overloaded on the gel, the monomorphic nature of the bands was clear (compare monomorphic bands produced by primer 8.10b between 250 bp and 400 bp in Fig 3, lanes, *G. soja* and *G. max* with Fig. 1). Some DAF profiles of *G. soja* and *G. max* were very similar, an example of which is shown with primer 8.6d (Fig. 2, lanes GS, GM). A single nucleotide change from GT AAC GCG (primer 8.6n; lanes GS, GM) to GT AAC GCC (primer 8.6d; lanes GS, GM) changed the DAF profile confirming the findings of Caetano-Anollés et al., [1991b] and Williams et al., [1990] that primer-template interaction is specific (at least as controlled by the 3' region of the primer).

Table 1 shows a catalogue for *G. soja* and *G. max* distinction using 8-mers varying from 50 to 100% GC. Forty polymorphisms were detected using 26 primers giving an average frequency of 1.5 polymorphisms per primer. Fifteen DAF markers belonged to Class I, twelve to Class II and thirteen to Class III. Primers consisting of different sequences but with fixed 62.5% GC content gave from 11 to 27 detectable DAF products per primer. The corresponding frequency of DAF markers detected ranged from 0 - 15%. Similarly, primers containing 75% GC gave from 16 to 28 detectable DAF products and a frequency of polymorphism ranging from 0 - 17.6%. A statistical test for independence of data in Table 1 (all 8.6 and 8.7 primers) showed that the total number of DAF products and the frequency of polymorphisms did not correlate with the GC content of the primer when tested at the 0.005 % level of confidence.

Inheritance of DAF markers

DAF polymorphisms of all three classes showed Mendelian inheritance in the F1 and F2 progeny. Chi square analysis of the segregation ratios of six DAF markers belonging to all three classes in the F2 population demonstrated true Mendelian inheritance (Table 2). The two phenotypic classes are 'presence' and 'absence' of the polymorphic band. Mendelian inheritance of class I and class III DAF markers obtained with primer 8.6n is shown in Fig. 2. The class I DAF marker at 240 bp was absent in *G. soja*, present in *G. max*, and segregated 30 : 6 in F2 progeny (only six F2 plants shown). The class III DAF marker at 205 bp was present in *G. soja*, absent in *G. max* and segregated 18 : 7 in the F2 population. A chi square analysis presented in Table 2 gave P values > 0.2 and > 0.7 for the 240 and 205 bp DAF markers respectively, which is in agreement with a 3:1 expected

Table 1: Primer catalog for *G. soja* and *G. max* distinction.

Primer code	Primer sequence (5'-3')	% GC	Number of amplification products[2] detected in		Number of AFLPs[5] belonging to				
			G. soja[1]	G. max[1]	Class I	Class II	Class III	Total	% AFLPs[3]
8.5a	AA TGC AGC	50.0	26	26	1	1	0	2	7.7
8.6a[4]	GA GCC TGT	62.5	22	20	2	0	0	2	9.1
8.6b	CC TGT GAG	62.5	24	24	0	0	0	0	0.0
8.6c	AA CGG GTG	62.5	14	13	1	0	0	1	7.1
8.6d	GT AAC GCC	62.5	22	22	0	0	0	0	0.0
8.6f	GA TCG CAG	62.5	16	14	1	0	1	2	12.5
8.6g	CT AAC CGC	62.5	21	21	0	0	0	0	0.0
8.6h	GA AAC GCC	62.5	20	20	0	1	1	2	10.0
8.6i	GT TAC GCC	62.5	16	19	1	0	2	3	15.8
8.6j	GT ATC GCC	62.5	20	22	1	1	0	2	9.1
8.6k	GT AAG GCC	62.5	27	26	1	0	2	3	11.1
8.6l	GT AAC CCC	62.5	20	21	1	0	2	3	14.3
8.6m	GT AAC GGC	62.5	12	11	0	1	0	1	8.3
8.6n	GT AAC GCG	62.5	21	21	1	0	1	2	13.3
8.7b	GC TGG TGG	75.0	26	24	1	0	1	2	7.7
8.7d	CC GAG CTG	75.0	19	22	1	1	0	2	9.1
8.7e	CC TGG TGG	75.0	20	21	0	1	0	1	4.8
8.7g	CC AGG TGG	75.0	18	18	0	0	0	0	0.0
8.7h	CC TCG TGG	75.0	25	27	2	0	0	2	11.1
8.7j	CC TGG AGG	75.0	27	28	1	1	1	3	10.7
8.7k	CC TGG TGC	75.0	17	16	0	2	1	3	17.6
8.7l	CC TGG TGC	75.0	27	27	0	0	0	0	0.0
8.9a	CG CGG CCA	87.5	16	16	0	0	0	0	0.0
8.9b	GG ACC CGC	87.5	19	17	0	1	1	2	10.5
8.10a	GG GGG GGG	100.0	0	0	0	0	0	0	0.0
8.10b	GC CCG CCC	100.0	14	12	0	2	0	2	4.3

[1] *G. max* = variety 'Bragg'nts382; *G. soja* = PI468.397; [2] as scored by visual comparison of bands between 100 and 700 bp on silver stained gels
[3] relative to genotype with the maximum number of amplification products and equivalent to number of AFLPs per total number of resolved bands
[4] amplification carried out with Stoffel fragment; [5] average number of AFLPs per primer is 1.5.

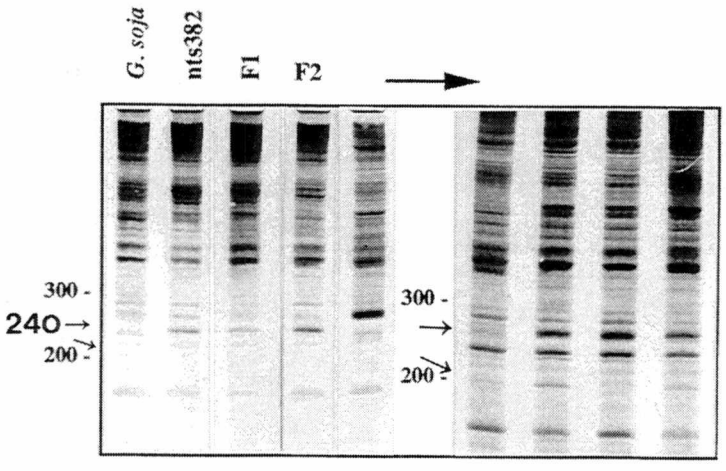


Figure 2: Mendelian inheritance of class I and class III DAF markers.
G. soja, *G. max*, F1 and F2 DNA (nts382 x *G. soja*) was amplified with primer 8.6n. Phenotypic classes being scored are presence (+) and absence (-) of band. The class I marker at 240 bp (marked with upper arrow) is absent in *G. soja*, present in *G. max* (nts382) and present in F1. Six representative F2 plants shown segregate +, -, -, +, +, + . The class III marker at 205 bp (lower arrow) is present in *G. soja*, absent in *G. max* present in F1 and segregates -, +, +, +, -, + in F2 progeny. Chi square analysis of the data is presented in Table 2.

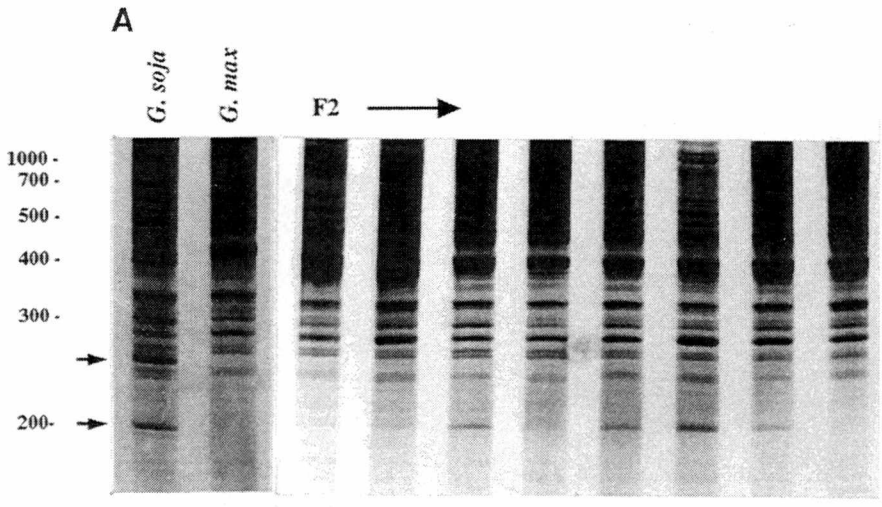


Figure 3: Mendelian inheritance of class II DAF markers.
G. soja, *G. max*, and F2 DNA (*G. soja* x nts382) was amplified with primer 8.10b. Phenotypic classes being scored are presence (+) and absence (-) of band. The 200 bp marker (arrow) is present in *G. soja* and absent in *G. max*. Eight representative F2 plants shown segregate -, -, +, -, +, +, +, - for the marker. Chi square analysis of the data is presented in Table 2.

Mendelian ratio. Mendelian inheritance was also observed with a class II DAF marker obtained with primer 8.10b (Fig. 3). A 200 bp DAF marker is present in *G. soja*, absent in *G. max* and segregated 20 : 8 in F2 progeny (eight F2 plants shown). Chi square analysis gave $P > 0.7$ which fits a 3:1 expected ratio (Table 2).

DAF markers with primer 8.6a (Table 2) were amplified using the Stoffel fragment of *Taq* polymerase. Two, class II DAF markers at 350 bp and 180 bp were present in *G. soja*, absent in *G. max* and segregated 28 : 8 and 30 : 6, respectively, in F2 progeny showing that both enzyme types produced molecular markers which segregated in a Mendelian fashion. Chi square analysis for a 3 : 1 fit gave P values > 0.7 and > 0.2 respectively.

Abnormal Inheritance Patterns

Abnormal (non-Mendelian) inheritance was observed for two out of eight DAF markers (Table 2). A 200 bp DAF marker (Fig. 4C) that was amplified with primer 8.7h, was present repeatedly in *G. soja* (the male parent), absent in *G. max*, and present in the F1 and all F2 plants tested. No segregation occurred in all F2 progeny tested (n = 32, 11 shown).

The second non-Mendelian DAF marker (250 bp) which was amplified using primer 8.10b, was repeatedly present in *G. soja*, absent in *G. max* but failed to be detected in the F1 (Fig. 4A) and F2 progeny (Fig. 3). Overloading the gels failed to reveal the marker band (Fig. 3, *G. soja* and *G. max*). Since F1 progeny contains equal amounts of *G. soja* and *G. max* nuclear DNA, a 1:1 mass ratio mix of *G. soja* : *G. max* DNA was amplified with primer 8.10b to test if competition between template sites was a factor in the non-Mendelian inheritance of this DAF marker. The synthetic F2 created the same genetic environment *in vitro* that was found in the F2 progeny. The DAF marker was, however, detected from the 1:1 mixture but did not amplify from F1 DNA (Fig. 4A, F1 and 1:1 *G. soja* : *G. max*). A quantitative comparison of lanes 1-4 in Fig 4A was done by laser densitometry. The scan (Fig. 4B) confirmed (1) the absence of the 250 bp DAF marker in both *G. max* and F1 progeny, since all other peaks representing monomorphic bands were consistently present, (2) the presence of the 250 bp DAF marker in a 1:1 mixture of *G. soja* and *G. max* DNA, and (3) the presence of the 200 bp DAF marker in F1 which was inherited as a Mendelian marker (Table 2).

Table 2: Chi square analysis of inheritance patterns of DAF markers in F2 plants from a cross between *G. max* line nts382 and *G. soja*.

DAF marker	Class[2]	Primer	Ratio[1] Expected	Ratio[1] Observed	χ^2	P
Mendelian segregation						
220 bp	I	8.7h	23 : 8	23 : 8	0.0	> 0.99
180 bp	II	8.6a	27 : 9	30 : 6	1.33	> 0.2
350 bp	II	8.6a	27 : 9	28 : 8	0.14	> 0.7
240 bp	I	8.6n	27 : 9	30 : 6	1.33	> 0.2
205 bp	III	8.6n	19 : 6	18 : 7	0.12	> 0.7
200 bp	I	8.10b	21 : 7	20 : 8	0.19	> 0.7
Abnormal segregation						
250 bp	II	8.10b	21 : 7	0 : 28	(maternal ?)	
200 bp	I	8.7h	23 : 8	31 : 0	(paternal ?)	

[1] Phenotypic classes = presence of band : absence of band
[2] Class I, II, and III refers to the visual classification of the band based on its intensity on DAF gels. See text for further explanations.

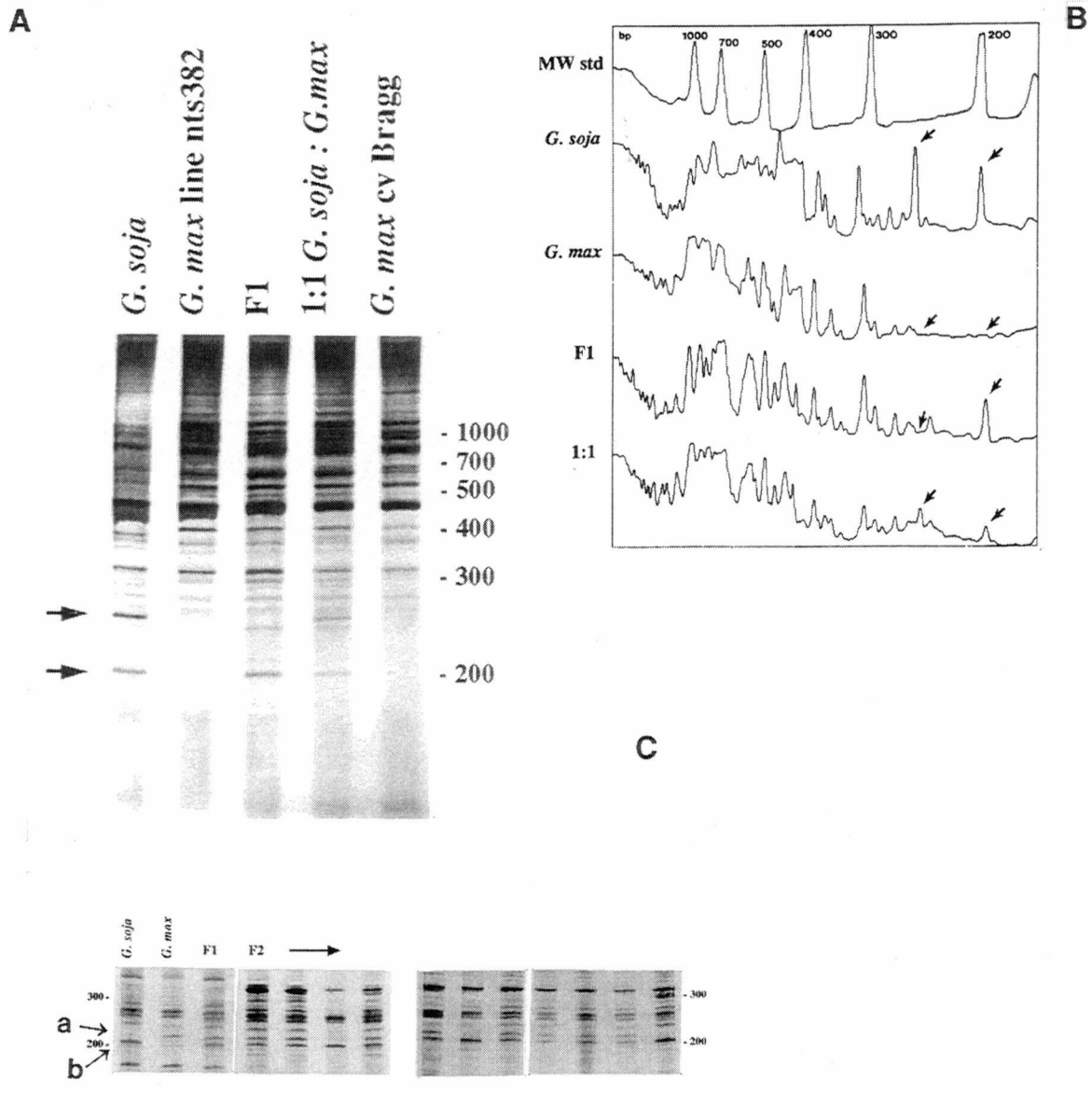


Figure 4: Non-Mendelian inheritance of DAF markers in soybean
(A) Maternal inheritance of DAF marker. Amplification of *G. soja* and *G. max* DNA using primer 8.10b shows a 250 bp AFLP (upper arrow) that was present in *G. soja*, absent in *G. max* line nts382, absent in F1 progeny, present in 1:1 ratio by mass of *G. soja* and *G. max* line nts382. The 200 bp marker (lower arrow) is present in *G. soja*, F1 and 1:1 mixture and shows Mendelian inheritance DAF profiles are identical in closely related lines, *G. max* line nts382 and *G. max* cv Bragg.
(B) Laser densitometric scan of DAF products from *G. soja, G. max*, F1 progeny, and 1:1 molar mixture of *G. soja* and *G. max* (1:1). The scan was produced using the first four lanes in fig. 7. The 200 bp AFLP is present in *G. soja*, F1 and 1:1 mixture (by mass) but is absent in *G. max*. The 250 bp AFLP is present in *G. soja* but absent in all others. The lane showing 1:1 mixture was stained lighter and shows smaller peaks.
(C) Paternal inheritance of DAF marker. Primer 8.7h (GCTCGTGG) amplified a 200 bp fragment (arrow) that was present in *G. soja*, absent in *G. max* and present in F2 progeny (11 F2 individuals shown).

Sensitivity of DAF analysis

DNA from *G. soja* and *G. max* was mixed at varying mass ratios in order to determine the sensitivity of DAF to amplify a polymorphic band from a mixture of template DNAs (Fig. 5). Reciprocal dilutions of 1:2, 1:6, 1:14, and 1:100 (latter not shown) of a mixture of *G. soja* and *G. max* DNA was amplified with primer 8.10b. The 200 (class I) and 250 (class II) bp polymorphic markers did not amplify if *G. soja* constituted less than 1:6 (17%) of the mixture.

While *G. soja* and *G. max* were suitable for studying the heritable characteristics of DAF markers and for revealing polymorphisms, the dominant nature of DAF markers makes them less suitable for mapping purposes in F2 populations. An alternative recombinant inbred mapping population was therefore considered for two main reasons. First, a map using RI lines existed [Lark et al., 1993] and second, the plant material and DNA was available. The lack of heterozygotes in the RI population makes it suitable for DAF markers wherein heterozygotes cannot be detected with existing methodologies.

Molecular mapping of DAF markers

The steps for mapping a DAF marker is discussed in Materials and Methods. Twenty-five oligonucleotide primers (8 mers) ranging in GC content between 60 to 100%, were used to screen parental DNA from soybean cultivars Minsoy and 'Noir1'; sixteen polymorphic markers were obtained (0.6 DAF marker/primer). Sixteen percent of the primers tested showed more than one polymorphic marker. Fifty percent of the primers that were used gave polymorphic markers in both populations. Except for one marker, the polymorphic markers for Minsoy and Noir 1 exhibited different molecular weights than for nts382 and *G. soja* PI468.397.

Table 3 shows mapping data for three DAF markers obtained by analyzing approximately 50 plants from recombinant inbred lines obtained from a cross between Minsoy and 'Noir1' [Lark et al., 1993]. The population is at F11 and contains mainly homozygotes (see part I). Three DAF markers DAF220, DAF233, and DAF234 mapped at distances ranging from 1.2 to 2.8 cM to existing markers and with LOD scores ranging between 7.4 and 12.5. The three markers mapped to three different linkage groups (Fig. 6). Marker DAF220 is located between pK644a and pA186 (2.1 and 2.2 cM respectively) on linkage group 4, while DAF233 was mapped between pA109b (5.1 cM) and BLT02 (2.8 cM) on

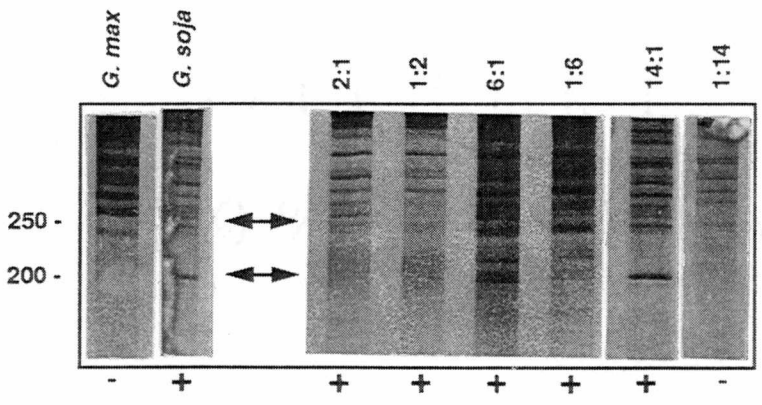


Figure 5: Sensitivity of DAF marker detection in a mixed-template reaction.
DNA from *G. soja* and *G. max* was mixed in different mass ratios of 2:1 (*soja:max*), 1:2, 6:1, 1:6, 1:14, 14:1 and amplified using primer 8.10b. The 250 bp and 200 bp markers (arrows) were absent in *G. max*, present in *G. soja* but did not amplify if *G. soja* constituted less than1:6 of the mixture. Note that the bands are present at 1:6 but absent at 1:14. The gel was overstained to detect any faint bands.

Table 3: Mapping data for three DAF markers on Utah-RIL map.

Marker	Primer[1]	Size[2]	LOD score[3]	cM distance[4]
DAF220	8.7h[5]	160	12.5	2.0
DAF233	8.7p[6]	185	7.4	2.8
DAF234	8.7p	160	11.7	1.2

[1] Refers to the primer that was used to generate the DAF marker.
[2] size in base pairs.
[3] Sample size: 58 RI lines for DAF220 and 47 RI lines for DAF233 and DAF234.
[4] Refers to cM distance from the nearest marker.
[5] The sequence of 8.7h is CC TCG TGG.
[6] The sequence of 8.7p is GCC GG CTA.

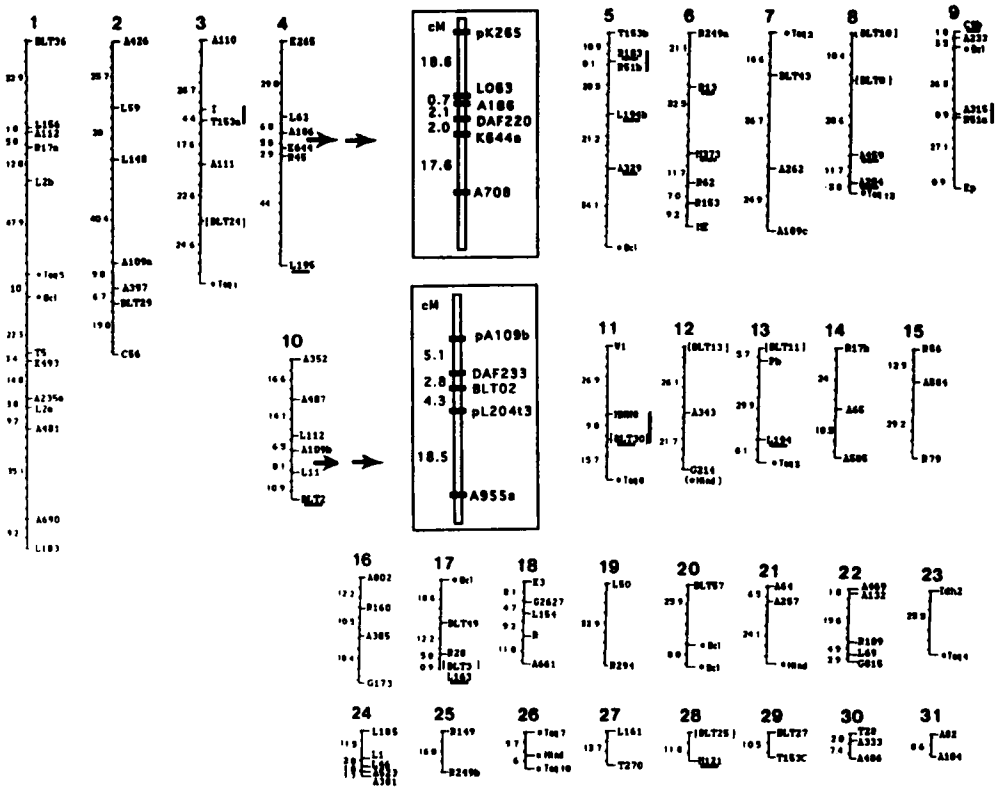


Figure 6: Intraspecific map of soybean[a] showing the position of mapped DAF markers among 31 linkage groups.
DAF233 and DAF 234 map to linkage group 4 and 10 respectively. Inset shows a close-up of the region containing the DAF markers.

[a] Taken with permission from Lark KG, Weismann JM, Matthews BF, Palmer R, Chase K, Macalma T: A genetic map of soybean (Glycine max L.) using an intraspecific cross of two cultivars: Minsoy and Noir 1. Theor Appl Genet 86: 901-906 (1993).

linkage group 10 (Fig. 6A). DAF234 is 1.2 cM from pG214H7 and a further 2.8 cM from pA401 on linkage group W1 (Fig. 7). The linkage group shown in Fig. 7 is not found on the map shown in Fig. 6A since the markers on LGW1 were mapped subsequent to the publication of the map and linkage groups have been redesignated using an alphanumeric system. DAF233, DAF220-A, and DAF220-B were the first DAF markers to be mapped in the soybean genome.

Although DAF233 and DAF234 were generated by the same primer 8.7p, they mapped to different linkage groups. A fourth 125 bp DAF marker generated by 8.6l did not show linkage presumably because of a small sample size and the incompleteness of the RIL map. Increasing the RIL sample size did increase the LOD score of this marker (Dr. K. G. Lark, pers. comm.)

The significance of the cluster of markers on linkage group W1 (containing DAF234) was fortuitously revealed when the early nodulin, ENOD2 was mapped. ENOD2 maps to the "I" locus and QTL controlling race 3 cyst nematode resistance *Rgh4* (F. Ghassemi pers. comm.). In addition, ENOD2 shows a high LOD score to the pod filling QTL on linkage group LGW1. This suggests possible epistastic interactions between the seed filling QTL on linkage group W1 and the "I" locus.

IV: DISCUSSION

This study showed that polymorphic markers generated by DAF can be used as genetic markers and that some markers show unexplained inheritance patterns. Mapping the first DAF markers onto the soybean map demonstrated the utility of DAF markers as an alternative source of molecular markers for genome mapping. The reproducibility of monomorphic bands between *G. max* and *G. soja*, shown in part III was extended in this study to show reproducibility between parents, F1 and with synthetic F1 (1:1 *G. max* : *G. soja*) by (1) the densitometric scans shown in Fig. 4B, where monomorphic bands were reproducible in four different genotypes, *G. soja*, *G. max*, F1, 1:1 mixture of *G. soja* and *G. max*, and (2) identical DAF patterns of two closely related lines of *G. max*, line nts382 and cv Bragg (Fig. 4A, *G. max* line nts382 and *G. max* cv Bragg). Using DAF, an average of 1.5 polymorphisms per primer was obtained between *G. max* and *G. soja* (and 0.6 per primer within *G. max* varieties). This represents a three-fold increase in frequency from previous studies by Williams et al., [1990] who showed that the frequency of RAPD polymorphisms using similar *Glycine* species as used in this study was 0.5 per primer.

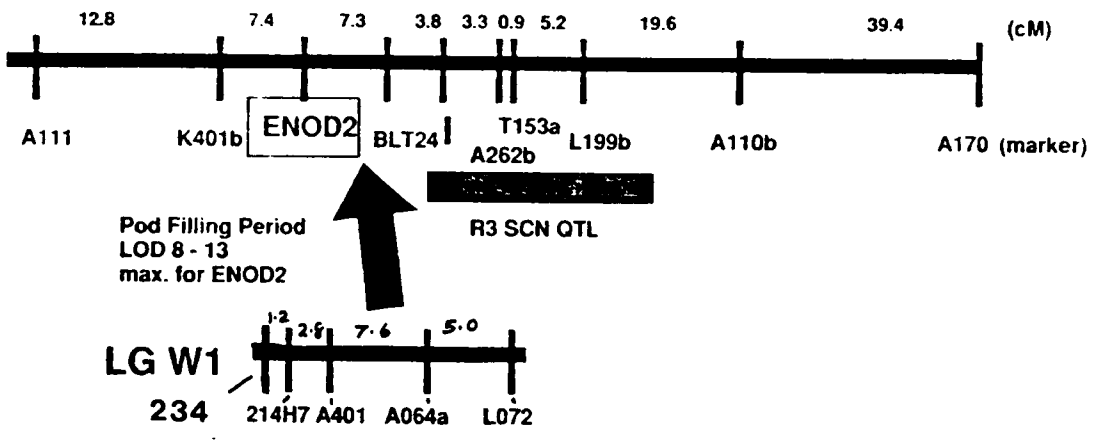


Figure 7: Map position of DAF234 on linkage group W1.
DAF234 maps to linkage group (LG) W1 which controls the QTL for pod filling period. ENOD2 showed a strong linkage to the "I" locus and the QTL for soybean cyst nematode (SCN) resistance to race 3 and shows linkage to LGW1 containing DAF234 (see text for explanations). I controls soybean seed coat color and represents a locus of chalcone synthase (Lila Vodkin, unpublished data)

Since the agarose-ethidium bromide detection produced about 5 to 15 bands, and DAF produced about thrice as many products, the observed difference in DAF marker frequency may stem from the detection method and not the shorter primer or altered amplification parameters [Gresshoff 1994].

The total number of DAF products shown in Table 1 ranged from 0 to 27 for primers between 50 and 100% GC, and did not show any correlation between the GC content and the number of amplification products. Theoretical calculations based on genome complexity, and GC content of primer and template shows that expected number of DAF products at 75% and 62.5% GC of the primer is 30 and 68 respectively [Caetano-Anollés 1994a]. Lowering the GC content did not produce the expected increase in the number of amplification products for DAF and some RAPD reactions [Caetano-Anollés 1994]. However a study that surveyed 480 decamers among distantly related taxa of flowering plants showed a positive correlation between the GC content of the primer and the ability and strength of amplification [Fritsch et al., 1993], confirming earlier findings with RAPD markers in soybean by Williams et al., [1990]. Mechanistic differences between RAPD and DAF reactions may be an explanation for the observed dependence of RAPD bands on primer GC content [Caetano-Anollés 1994a]. Analysis of DAF markers showed an inability to detect quantitative differences in DAF markers between the heterozygote F1 and parent (Fig. 5). To circumvent this limitation with F2 populations, a soybean RIL mapping population was used. Three DAF markers were mapped in the soybean genome and demonstrated good linkage with LOD scores > 4 (Table 3). All three markers mapped to different linkage groups even though two of the markers were generated by the same primer showing that each marker is capable of defining a different region on the map. The fortuitous mapping of DAF234 to linkage group W1 that is associated with the QTL for seed filling period demonstrates that not only can DAF markers be used as genetic markers, they map to important regions of the genome. A fourth DAF marker did not show linkage to any existing markers on the map but showed an increased LOD score when the progeny sample size was increased. The data for the unlinked marker may reveal linkage to as yet unmapped markers or may be important in linking two different linkage groups in future.

Under optimal conditions, successful DAF amplification depends on the presence of sites in the genome that are complimentary to the arbitrary primer used. Therefore monomorphic bands in a DAF profile could conceivably arise from two different sequences (unlinked) in the genome which comigrate on the gel. Comparing RAPD maps for two sets of maize

recombinant inbred lines, Johns [1992] found that instances of two unlinked bands having the same mobility is rare. He also showed that RAPD polymorphisms in one cross are not polymorphic in another and that bands of similar sizes (on agarose gels) in the two crosses usually map to two different unlinked locations. This was confirmed in the present study, as a polymorphic band generated by primer 8.7h in the *G. max* x *G. soja* cross was not polymorphic in the intraspecific cross between Minsoy and Noir 1.

Six of eight DAF markers belonging to three intensity classes were inherited in the F2 progeny of the *G. max* x *G. soja* cross at the expected 3:1 Mendelian ratio. In addition to producing normal, heritable DAF markers, primers 8.10b and 8.7h amplified polymorphisms that showed non-Mendelian inheritance. The ability of one primer to generate a Mendelian as well as a non-Mendelian marker is significant and eliminates artifactual explanations. A 250 bp DAF marker amplified with primer 8.10b failed to be inherited in F1 and F2 progeny. The absence of the 250 bp DAF marker was surprising since soybean is an inbred plant and all genomic regions are assumed to be homozygous. Varieties are generally the product of recurrent selection and single seed descent. Nevertheless, the possibility of inheritance of a null allele of a heterozygous locus in the *G. soja* individual used to make the F1 and subsequent F2 generation was considered since the DAF marker was absent in all 28 F2 plants tested (Table 2). However, DNA from eight individual *G. soja* plants showed that the 250 bp DAF marker was present in all *G. soja* plants (data not shown). Assuming only 10 generations of selfing for *G. soja*, the expected ratio of presence of band : absence of band among individual *G. soja* plants is 1:1. The observed ratio of 8:0 (presence of band : absence of band) did not fit the expected 1:1 ratio ($P<0.01$). The *G. soja* material therefore was considered to be homozygous. Similar results by Gresshoff and MacKenzie [1994] indicated that individual Bragg plants gave identical amplification profiles, thus genetic heterogeneity of either parent was not an explanation.

Additional control experiments eliminated alternative explanations for the disappearance of the 250 bp DAF marker. First, the absence of the 250 bp DAF marker was confirmed in F1 progeny from a separate cross between *G. soja* and *G. max* line Bragg nod49 (almost identical to line nts382), thereby eliminating the possibility that this DAF marker arose from a particular cross (J. Blauenfeldt pers. comm.). Second, given the reproducibility of DAF markers between gels and amplifications, it seems unlikely to be an amplification or staining artifact. Third, mixtures of *G. soja* and *G. max* showed that the 250 bp band was synthesized, when mixed at a 1:1 mass ratio suggesting that competition between templates

was not a factor. Since, the 250 bp polymorphic band was not synthesized, if *G. soja* constituted less than 17% of the mixture and *G. soja* was used as the male parent, it is likely that the 250 bp DAF marker originated from *G. soja* organelle DNA which constitutes 10 to 20% of total cellular DNA [Grierson and Covey 1988]. Since soybean organelles are presumed to be inherited maternally, the DAF marker did not appear in F1 or F2 progeny. This particular DAF marker, therefore, shows potential for identifying the *G. soja* cytoplasm. The reciprocal cross, *G. max* x *G. soja* was not available because of difficulty in the cross-pollination. Alternatively, it is feasible that the template for this DAF marker is not passed to the next generation and thus can be considered to be epigenetic.The absence of the 250 bp marker in F1 and F2 progeny raises concerns about the use of these markers for paternity analysis. Although there have not been any published reports regarding the absence of RAPD bands in F1 and F2 progeny, non-parental bands in the progeny DNA has been reported in baboon, human CEPH pedigrees [Reidy et al., 1992], beetles (*Nicrophorus tomentosus*: Silphidae) [Scott et al., 1992] and bees [Hunt and Page 1992]. The mechanisms for producing non-parental bands include "jumping PCR" [Pääbo et al., 1990] and the formation of heteroduplex molecules of altered mobility (see below). Jumping PCR is a phenomenon exhibited by DNA polymerase which "jumps" between two molecules that show regions of extensive homology to produce a chimeric molecule. Since the "jump" occurs when the polymerase encounters the end of a template molecule, this phenomenon occurs at a high frequency with damaged template DNA.

A high frequency of non-parental bands was reported from primates (an average of 4.4 and 2.7 bands generated by five primers) [Reidy et al., 1992]. In case of bees, in the absence of DNA polymerase, amplification resulted in the formation of 'diploid specific' bands (10% frequency) suggesting the formation of sequence heteroduplexes between allelic PCR products [Hunt and Page 1992]. A non-parental band was detected in F1 progeny molecules of the flax rust fungus, *Melamspora lini* (Ehrenb.) Lev., that was not amplified from either parent [Ayliffe et al., 1994]. Of the 640 bands generated by 160 RAPD primers, one heteroduplex molecule (frequency of 0.16%) was formed between 2 allelic sequences of different sizes and which resulted in a band of altered mobility in a heterozygote F2. The altered mobility was a result of the presence of repeated sequences at the ends of the molecule that caused a change in the structure of the molecule. In the beetle study, out of 1450 larvae-reactions only 5 (0.3%) of the reactions produced non-parental bands. The authors postulate that the bands could have been generated from contaminating DNA originating from bacteria, parasites, fathers not represented in the sample population

or other unknown reasons [Scott et al., 1992]. Non-parental bands may make APT markers unsuitable for certain problems in evolutionary biology. However, they can still be useful for genotyping and paternity tests. Important considerations for these applications are reproducibility, the use of several diagnostic bands and using band presence to be diagnostic whereas band absence as being uninformative.

A 200 bp DAF marker amplified with 8.7h showed no segregation in F2 plants. Since the 220 bp DAF marker with the same primer on the same gels gave Mendelian segregation, and since *G. soja* was the male parent, explanations involving traditional cytoplasmic inheritance are not possible. A 31:0 segregation ratio could be expected if the 200 bp DAF marker arises from multiple loci that are present on separate chromosomes in *G. soja*. Independent segregation of the multiple loci in the F2 progeny could result in the inheritance of at least one copy by every F2 progeny plant in this subsample. Alternately, this marker may define a cytoplasmic replicon in *G. soja* that is inherited in a yet unreported paternal fashion. It is also possible that the marker originated from a region of the genome that is essential for seed viability in F2 and seeds that do not show the DAF marker do not exist. This implies that this region is essential in the F2 progeny but not in *G. max* since the band is not amplified in this parent.

Eleven percent of the RAPD markers in *Stylosanthes* showed distorted segregation, and 13% showed no segregation in the F2 progeny [Kazan 1993]. One marker was of maternal chloroplast origin as detected by "chloroplast enriched" DNA. In corn, Heun and Helentjaris [1993] found that 10% of the RAPD markers show non-Mendelian inheritance when tested on 16 F1 hybrids. As discussed in part III, amplification reaction conditions, scoring errors or genetically-based errors can be considered to be the sources of these non-Mendelian segregating bands. Weeden et al., [1993] determined that in pea, an error rate of approximately 4% was due to reaction conditions. Analysis of double recombinants in apple showed a significant increase of 2 to 7% from expected values indicating errors in the amplification process itself. These suggests that the genetic background of the marker can be a factor to the successful amplification of a polymorphic band and that errors due to non-optimal reaction conditions cannot be a sole factor in abnormal inheritance. Distorted F2 ratios have been observed for both isozyme and RFLP markers in soybean and other crops [Edwards et al., 1987, Slocum et al., 1990, Tulseiram et al., 1991].

In summary, about 25% of DAF markers show non-Mendelian inheritance that can be attributed to either an organelle origin for these markers or to altered genetic background.

PART V*

USE OF DNA AMPLIFICATION FINGERPRINTING FOR GENOTYPE DISTINCTION AND PEDIGREE ASSESSMENT OF SOYBEAN CULTIVARS

*Prabhu RR, Jessen H, Webb D, Luk L, Smith S, Gresshoff, PM: Crop Science (in preparation)

I: INTRODUCTION

In plant breeding, information on genetic relationships within and between species is used for organizing germplasm, selection of parents and limiting the number of accessions required to ensure that a wide range of genetic variability is being sampled. Genetic distance estimates are also used for identifying heterotic groups for crops. Seed companies have more than a casual interest in evaluating germplasm and identification of cultivars. The ability to accurately identify cultivars or F1 hybrids is crucial in order to maintain control over marketing, propagation, quality, and patenting rights. Protection of varieties through fingerprinting techniques must meet certain criteria. First, the fingerprint should not vary with the environment. Phenotypic identification based on scoring morphological traits in the field are expensive, labor intensive and environmentally plastic. Moreover, morphological traits often cannot be described unambiguously and therefore are limited in their use as for identification of cultivar. Second, the identification technique must have high discriminatory power to distinguish isogenic or inbred lines. Biochemical techniques such as electrophoresis of isozymes, and HPLC analysis of seed storage proteins [Yadava et al., 1979] have been used successfully for profiling but these strategies are limited in the amount of polymorphisms that are detected. Third, information on genetic distances must be extractable from the fingerprinting data. Fourth, the technology for fingerprinting should be accessible at a level that permits widespread use.

These criteria are met by profiling with arbitrary primer technologies and are therefore gaining widespread usage for cultivar identification and genotype relationships (see part I). The objective of this study was to use DAF to distinguish ten anonymous soybean (*Glycine max*) genotypes supplied as seed material from Pioneer Hi-Bred International, Inc., (Drs. S. Smith and D. Webb), and use the data to reveal relationships between the genotypes. The DAF dendrogram is compared to dendrograms generated by RFLP and pedigree data for the same ten varieties. The purpose of this study was to use DAF to distinguish and predict the relationships among an unknown set of soybean varieties.

II: MATERIALS AND METHODS

For the unknown soybean varieties from Pioneer Hi-bred, methods for isolation, amplification and separation were identical to those described in parts III and IV. Data analysis for cluster analysis is described in part VIII, Materials and Methods.

II: RESULTS

Distinction of ten soybean cultivars using seven DAF primers

Those primers which gave informative profiles for soybean (part IV and [Prabhu and Gresshoff 1994]) were used to select the primers used in this study. Table 1 shows resolution of soybean samples with seven DAF primers including one multiplex analysis, i.e., two primer analysis [Caetano-Anollés et al., 1991b]. Eighteen polymorphic bands (2.6 per primer) generated fingerprints that distinguished all samples except samples 1 and 2 (Table 4). The ten supplied samples varied in DAF marker differences. Per primer, between 1 and 4 polymorphisms ranging from 125 to 700 bp were detected of which 78% were in the 125-300 bp range. Primer 8.7h gave a weak polymorphic band with sample 3, which made it difficult to score the band consistently (Table 1). Hence a 0/1 designation was given, but for scoring purposes the data point was taken as a 1 (presence). Table 2 demonstrates how two primers, 8.7h and multiplex (8.7h + 8.5a) were sufficient to reveal informative differences between the samples (except for samples 1 and 2). Primer 8.7 h alone distinguished the soybean samples into six classes and the combined haplotype produced by 8.7h and multiplex (mp) distinguished the samples into nine classes. Additional polymorphic bands of identical in molecular weight as seen with 8.7h were produced with multiplex analysis (mp) but these were not included in the analysis.

Information on cultivar/line identity and RFLP data was provided to us after the DAF analysis was completed. Table 3 shows the identity and pedigrees of the ten soybean samples. Three lines, P9392, P9341, and P9171 are half sibs and have a common parent, A3127. Two pairs of lines are related by backcrossing; Archer, BSR101 and Hartwig, Forrest. Archer is backcrossed five times to the recurrent parent, BSR101 and consequently the two genotypes share 98.4% of their genomes. Hartwig is backcrossed twice to the recurrent parent Forrest, and therefore shares 87.5% of its genome with Forrest. P9593 and S8330 are the Southern most varieties. *Glycine soja*, the wild progenitor is genotypically most distant from the other nine cultivated varieties. Since different pairs of cultivars are related to different degrees, these soybean samples provide a good indicator of the sensitivity of DAF. The cultivated soybeans in this sample set represent agronomically important cultivars. Hartwig is the first cultivar with resistance to all races of soybean cyst nematode (SCN). It is also resistant to root knot nematode, reniform nematode and shows a high level of resistance to Sudden Death Syndrome.

Table 1: DAF-generated polymorphisms distinguish ten soybean genotypes.

primer	8.7h				mp		8.6n				8.6k			8.6a	8.10b	8.6j		
poly. band (bp)	200	210	250	300	125	300	180	200	250	290	225	380[a]	460	150	500	220	230	700
Genotype																		
1	0	1	0	0	1	0	1	1	1	1	1	1	1	0	*	1	0	1
2	0	1	0	0	1	0	*	*	*	*	1	1	1	0	*	1	0	1
3	1	1	0/1	1	0	0	1	1	1	1	0	0	0	0	0	0	0	0
4	1	1	0	1	1	0	1	0	1	1	0	0	0	1	*	0	0	0
5	1	0	0	1	1	0	0	1	1	0	0	1	0	1	1	1	1	0
6	0	1	1	1	1	1	1	1	0	1	0	1	0	0	0	1	0	1
7	0	1	1	1	0	0	1	1	1	1	0	1	0	0	*	1	0	1
8	0	1	1	0	1	1	1	1	0	1	0	0	0	0	0	1	1	1
9	1	1	1	1	0	1	1	0	1	1	1	1	1	0	0	0	0	1
10	1	1	0	1	0	1	1	0	1	1	1	1	1	0	0	0	0	0

1= presence of band; 0=absence of band; 0/1=weak presence of band in some gels. mp=multiplex DAF(8.5a+8.7h).
Multiplex also showed 200 and 210 bp generated by 8.7h (not shown). bp=base pairs. *=missing data;
For 380[a], 1=doublet of bands with a large gap and 0=doublet at that size with small gap

Table 2: Distinction of nine soybean genotypes using two DAF primers.

	8.7h					mp[3]			combined
gen[1]	200[2]	210	250	300	haplotype	125	300	haplotype	haplotype
1	0	1	0	0	a	1	0	A	aA
2	0	1	0	0	a	1	0	A	aA
3	1	1	0/1	1	b or c	0	0	B	b/cB
4	1	1	0	1	c	1	0	A	cA
5	1	0	0	1	d	1	0	A	dA
6	0	1	1	1	e	1	1	C	eC
7	0	1	1	1	e	0	0	B	eB
8	0	1	1	0	f	1	1	C	fC
9	1	1	1	1	b	0	1	D	bD
10	1	1	0	1	c	0	1	D	cD
total no. of classes					6			4	9

[1] refers to unknown genotypes, same as in table 1
[2] refers to base pairs (size of the marker)
[3] mp refers to multiplex primers or two different primers, 8.5a and 8.7h. The sequence of 8.7h is CC TCG TGG and the sequence of 8.5a is AA TGC AGC.

Table 3: Identity and relationships of soybean genotypes.

Genotype	Variety	Pedigree	Relationship
1	Archer	5 backcrosses with Williams 82 as donor and BSR101 as recurrent parents	
			Archer and BSR101 share 98.4% of their genomes
2	BSR101	L69U40-16-4 x A76-304020	
3	Forrest	Dyer x Bragg	
			Forrest and Hartwig share 87.5% of their genomes
4	Hartwig	2 backcrosses with PI 437.654 as donor and Forrest as recurrent parents	
5	PI 437.654[c]	*Glycine soja*	
6	P9171[a]	P9271 x A3127	
7	P9341	MO304-21BR x A3127	P9171, P9341, and P9392 are half sibs with A3127 as the common parent
8	P9392	A3127 x Gnome	
9	P9593	(P5482 x Centennial) x P9561	
10	S8330	Pedigree not avilable[b]	

[a] P indicates Pioneer Hi-bred variety

[b] variety from Northrup King, pedigree not available

[c] has resistance to all the known races of soybean cyst nematode

Forrest, a widely sold cultivar, is resistant to races 1 and 3 of SCN, root knot nematode and has excellent shatter resistance. P9593 is resistant to frogeye leaf spot, SCN race 3, and shows high yields. P9392 is susceptible to SCN and *Phytophthora* rot but shows good shatter rating.

DAF profiles obtained with 8.6n and 8.7h are shown in Fig. 1. Primer 8.6n showed 4 polymorphic bands and 22 monomorphic bands between 100 and 1000 bp. The fingerprint of Forrest (1111, see Table 1) was different from that of Hartwig (1011) by only one polymorphism, the 180 bp marker. The 180 bp fragment distinguished Forrest from Bragg and suggests that it originated in Dyer, a parent of Forrest. *G. soja* (0110), however, differed from Forrest and Hartwig by two polymorphic markers at 290 bp and 180 bp. The profiles of Archer and BSR101 showed no polymorphism. With primer 8.7h, the fingerprint (1001) of *G. soja* PI 437.654 differed from P9341 (0111), P9392 ((0110) and P9593 (1111), by markers at 200, 210 and 250 bp. The marker at 300 bp distinguished P9392 from P9593. Both primers 8.6n and 8.7h suggested that sample 5 (*G. soja*) was very distant from the other varieties.

Pedigree assessment

Fig. 2 shows three dendrograms for the ten soybean genotypes based on cluster analysis of pedigree, RFLP and DAF data. The overall pattern of relationships using DAF and RFLP data is similar to pedigree data the actual distances being slightly different among the three dendrograms. The DAF and RFLP data are in close agreement to the pedigree data in identifying major groups, group I (*G. soja* PI 437.654), group II (P-9392, 9171, 9341, BSR101, Archer) and group III (Forrest, Hartwig, P9593, S8330). As predicted by the data in Table 1, *G. soja* (sample 5) is the most distant from the other genotypes. The relative positions of individual members within the sub-groups is similar in the three dendrograms with some minor differences. Within the primary group, S8330 is separate from the Forrest/Hartwig cluster and P9593 in the pedigree and RFLP data. The DAF dendrogram shows a different relationship between S8330 and the other cultivars within the sub-group. However the distance matrix (data not shown) shows that in the DAF data, S8330 is at the same distance from Forrest and Hartwig as in the RFLP data. Its closeness to P9593 (0.09 units) results in a different distance relationship. The other difference in the DAF dendrogram is the non-distinction of Archer and BSR101. One presumes that the use of two or three primers could resolve these genotypes.

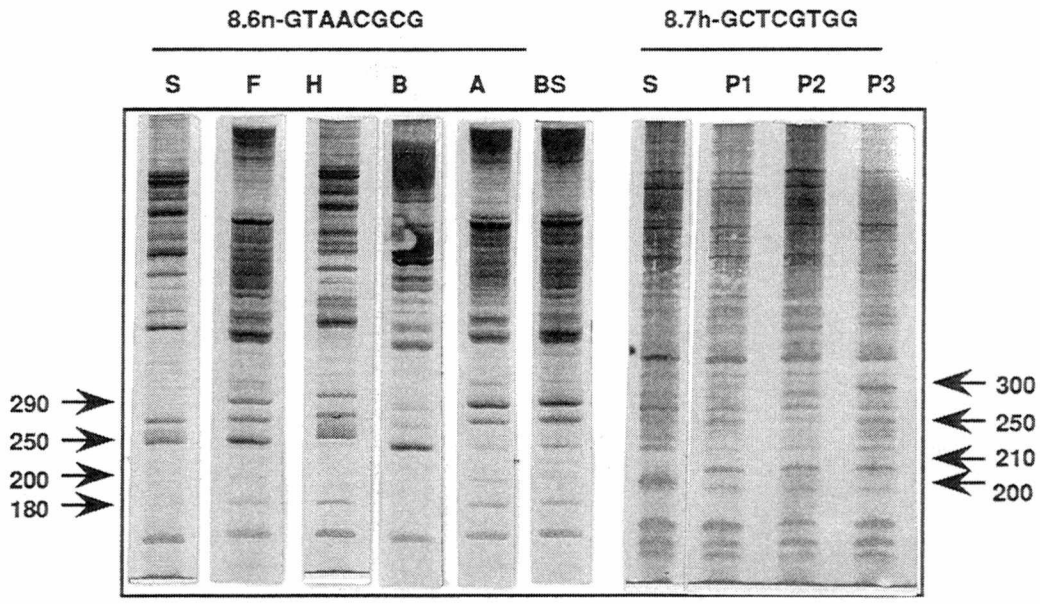


Figure 1: Distinguishing DAF fingerprints of 8 soybean genotypes.
Primer 8.6n showed four polymorphic fragments (arrows) of which the 290 and 180 bp markers distinguished *G. soja* (S) from Forrest (F) and Hartwig (H), the 200 bp marker distinguished between Forrest and Hartwig, two closely related lines, and the 250 bp marker was common to all three genotypes. The 180 bp fragment also distinguished Forrest from Bragg (B), a parent of Forrest and suggests that it originated in Dyer, the other parent of Forrest . The profiles of two closely related lines, Archer (A) and BSR101 (BS) are very similar.
Primer 8.7h showed four polymorphic fragments (arrows), of which the 300 bp fragment distinguished between P9392 (P2) and P9593 (P1). Markers at 200, 210 and 250 bp distinguished between *G. soja* (S) and P1 (P9341) but were common between P1, P2, and P3 which are half-sibs.

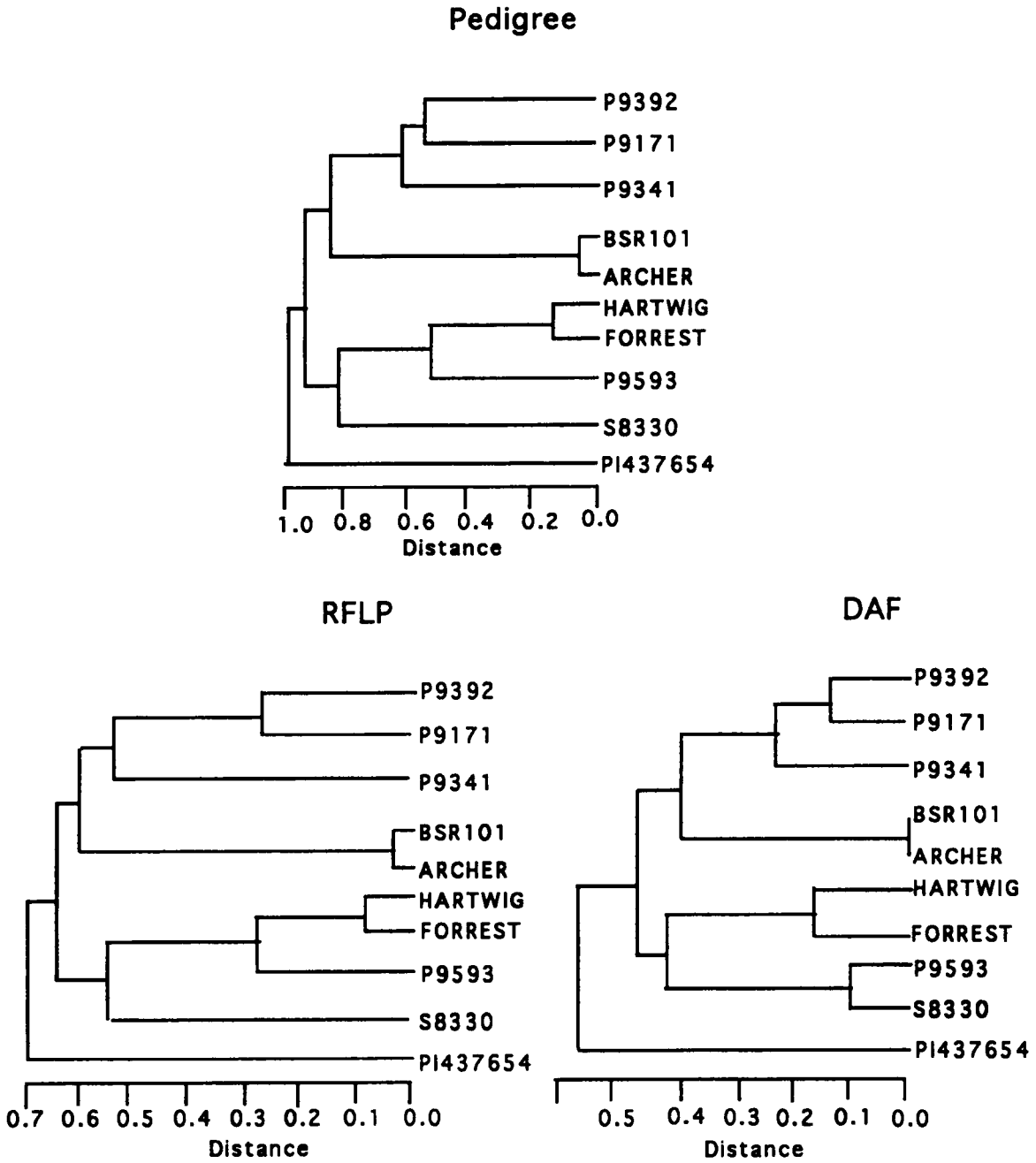


Figure 2: Relationship between 10 soybean genotypes revealed by cluster analysis of data based on pedigree, 53 RFLP probes, and 7 DAF primers

Discriminatory power of DAF primers

Seven DAF primers provided different levels of resolution between and within each major group (Table 4). Primers 8.7h and 8.6j were the most informative with the ability to distinguish at all levels; between major groups, primary groupings within the group and individual lines within each cluster. Six out of seven primers distinguished between closely related individuals within a cluster.

IV: DISCUSSION

It was possible to use DAF to resolve the ten samples supplied by Pioneer into nine genotypic classes. Archer and BSR101 could not be distinguished by DAF. The efficiency of DAF was demonstrated by the fact that only seven primers distinguished nine soybean genotypes with a high frequency of polymorphisms per primer (2.6). DAF primers varied in their discriminatory power and showed from one to four polymorphisms per primer. Primer 8.7h and 8.6n gave the highest frequency of polymorphisms, and 8.7h was valuable in multiplex analysis with primer 8.5a. Additional primers will be required to distinguish Archer and BSR101. Only two primers 8.7h and 8.5a (mp) were able to distinguish nine genotypes in this study. With a growing database of DAF markers it should be possible to identify a few primers that distinguish a large number of cultivars. This makes DAF a valuable and feasible technique for cultivar identification.

The utility of DAF to estimate genetic relationships between soybean genotypes was demonstrated in this study. The dendrograms generated by DAF data was similar to those based on pedigree and RFLP data. Both DAF and RFLP data have similar sensitivity for detecting the relationships between major groups and between individuals within sub-groups. Although DAF and RFLP produced equivalent data, 53 probes were required for RFLP whereas only 7 primers (including a multiplex) were required for DAF to generate the dendrograms. In terms of resources and ease of generating data, DAF can be considered to have an advantage over RFLP analysis (assuming that all 53 probes were required to distinguish 10 genotypes).

Although a dendrogram showing pedigree of genotypes is a useful standard for comparing those dendrograms generated by other marker systems, genetic distances in pedigrees must be interpreted with caution as they are likely to have hidden assumptions. For example, of the ten genotypes sampled in this study, nine are cultivated soybeans and *G. soja* is the

Table 4: Frequency of polymorphisms and resolution obtained with seven DAF primers.

primer	sequence (5' to 3')	number of polymorphisms	Resolution[a]
8.6a	GAG CCT GT	1	3
8.6j	GTA TCG CC	3	1, 2, 3
8.6k	GTA AGG CC	3	3
8.6n	GTA ACG CG	4	1, 3
8.7h	CCT CGT GG	4	1, 2, 3
8.7h + 8.5a multiplex	AGC TTG TC[b]	2	3
8.10b	GCC CGC CC	1	1
Total		18	

[a] 1: distinguishes primary clusters (wild type from cultivated varieties)
2: distinguishes within primary groupings
3: distinguishes individual lines within a cluster

[b]sequence for primer 8.5a only

wild-type progenitor. Assuming that correct information regarding the parents and grandparents is available, and estimates of genome similarities can be made, the degree of relationship between the original parents is not known. Also, as was done in this case, after going back several generations, an assumption is made that *G. soja*, the wild progenitor, is completely unrelated to the cultivated soybeans. All genetic distances are calculated relative to *G. soja* which is assigned a genetic distance of one (Fig. 2). The point being made is that while pedigree dendrograms give the overall relationships between the genotypes, the distances shown in the dendrogram need not reflect actual sequence similarities or dissimilarities between the genotypes. Since the primers or probes actually sub-sample genomic sequences, DAF and RFLP markers can be considered to be better estimators of genetic distances.

The difference in the actual genetic distance generated by RFLP and DAF compared to that predicted by pedigree data can be attributed to the type of genomic regions sampled by RFLP and DAF. RFLP clones are mostly cDNA clones or random *Pst*I clones and therefore RFLP polymorphisms are a reflection of variation in low copy DNA (coding regions). In addition if a sub-set of these *Pst*I clones are used as was done in soybean, this could result in a bias in the kind of regions sampled in the genome. Although no published information exists on the genomic equivalent of DAF fragments (see part 6) hybridization experiments with RAPD fragments indicate that RAPD fragments do not exhibit preferential amplification for any particular part of the genome [Williams et al., 1991]. Also the total number of DAF polymorphism and the number of DAF polymorphism detected does not correlate to the GC content of the primer further suggesting that DAF fragments represent an average sampling of the whole genome. It is interesting that inspite of approximately 20 fold difference in the length of the probe between RFLP (assuming 300 bp) and DAF (8 x 2 = 16 bp), both methods generated equivalent dendrograms. The closeness of P9593 and S8330 shown by DAF was at variance with that shown by RFLP and pedigree data. It is intriguing to consider that DAF is actually more sensitive for revealing differences at the sequence level for all the reasons stated above.

Unlike RFLP markers, APT markers are considered dominant, i.e., the presence of bands of equal molecular weight (monomorphic) in both genotypes does not necessarily indicate that they originated from allelic equivalents in the two genomes. In *Stylosanthes*, a forage legume, 200 amplification produced by 22 primers were used to distinguish 4 species. The level of polymorphism was low (0-16%) within species, higher (28-46%) among species,

but very low (0-2%) among individuals within each cultivar. The genetic relatedness was in good agreement with morphological, cytological, and isoenzyme studies. A study with cruciferous species showed that RAPD markers were very similar to RFLP markers for estimating intra-specific genetic relationships [Thormann et al., 1992]. However, the two marker types gave different results for inter-specific relationships. In both dendrograms based on RFLP data, *B. napus* was grouped closer to *B. napa* than to *B. oleraceae*. With RAPD based dendrograms, however, *B. napus* was closer to *B. oleraceae*. Further experiments showed that the difference in the RAPD dendrograms did not arise from the contaminating presence of RAPD bands arising from organelle DNA. Analysis of inter-specific monomorphic RAPD bands by Southern hybridization showed that RAPD bands with similar mobility may have varying degrees of homology and cause differences in the RFLP and RAPD dendrograms [Thormann et al., 1992]. This may also explain the closeness between P9593 and S8330 in the DAF data. The similarity in pedigree, RFLP and DAF dendrograms in this study is encouraging in terms of conservation of genetic localization of amplified products between two genotypes. However, more data describing the genetic basis of DAF polymorphisms across genotypes of a species are required to allow DAF-based dendrograms to be considered a good alternative to existing methodology.

In conclusion, DAF has the capability to meet all the criteria for varietal profiling and determining relationships; 1) reproducibility, 2) objective scoring of bands to allow statistically significant databases for valid comparisons of profiles across laboratories (DAF archivable gels make this easier), and 3) ability to generate identical equivalently meaningful data for distance calculations.

PART VI

GENOMIC BASIS OF DNA AMPLIFICATION MARKERS

Polymorphisms detected by arbitrary primer technologies are mainly dominant markers and can be produced by insertions or deletions in the DNA sequence or by changes in the primer binding site. However, if an insertion or deletion in the region amplified by the primer is within the amplification limit, it can result in a marker that is co-dominant. Characterizing the sequence amplified by the primer therefore is informative in terms of the allelic nature of the marker. More importantly, the sequence information can be used to convert the DAF marker into a SCAR [Paran and Michelmore 1993] which serves as a permanent mapped landmark in the genome. To convert a DAF marker into an RFLP marker it must first be cloned to obtain sufficient quantities for sequencing or manipulation. Although it is possible to radiolabel the band [Weaver et al., 1994] during amplification for use as a probe for Southern hybridization (described in Materials and Methods), this only applies to the situation where a single band of a single species is obtained upon amplification. Since one cannot be absolutely certain that there is a single species in the amplified probe, cloning is the best possible alternative. For the same reasons direct sequencing of the purified band from the gel must be approached with caution although methods for direct sequencing of PCR products are available [Liu et al., 1993, Cao et al., 1993].

Several methods have been used for cloning PCR products. These are briefly outlined below. (1) T/A methods use vectors with an extra thymidine (T overhang) to ligate the PCR product with an extra adenosine residue at its 3' end which is added by *Taq* polymerase [Ichihara and Kurosawa 1993; Kovalic et al., 1991]. (2) Blunt end cloning methods in which the extra adenosine residue is removed from the PCR product by Klenow or T4 DNA polymerase and the PCR fragment is ligated to a blunt-ended vector [Costa and Weiner 1994]. (3) Methods that incorporate restriction sites in the design of PCR primers and can be ligated to vectors cut with the same restriction enzyme [Lorens 1991]. (4) Methods that utilize exonuclease III treatment of the PCR fragment and vector to induce a stepwise release of nucleotides from the 3' end. The 5' end of the PCR primers have the same bases as those near the restriction site of the vector allowing for ligation between fragment and vector [Hsiao 1993]. (5) Methods that incorporate uracil into PCR primers and are subsequently cleaved with uracil-*N*-glycosylase to expose cloning sites [Nisson et al., 1991].

The sequence of arbitrary primers by their very nature cannot be manipulated, therefore, only the first two methods that do not require sequence manipulation are amenable for use with DAF markers. With the final objective of converting the mapped DAF markers into RFLP probes and using their sequences to synthesize SCAR primers, the objective of this section is to develop an easy and reliable method for cloning APT markers and use the cloned and sequenced markers for genomic analysis of DAF polymorphisms. Since most of the published literature deals with cloning PCR products, and not APT markers, a cloning protocol is described for cloning DAF markers. The protocol was developed with the objective of having a reliable, relatively fast, and general cloning method that did not require the purchase of expensive kits or the need for special cloning vectors. Some problems encountered when cloning polymorphic fragments are also presented. DAF233 and DAF220 were chosen as representative markers for conversion into RFLP probes since these markers had been mapped (see part IV) and were the first markers to be mapped in the soybean genome.

II: MATERIALS AND METHODS

The polymorphic bands were first purified from the DAF gel and cloned in order to convert them into RFLP probes. For cloning DAF fragments, the polymorphic band was first purified from the gel, and then cloned into the pBluescript (pBS^+) plasmid. The pBS^+ vector along with the ampicillin resistance gene, contains the *lacZ* gene coding for β-galactosidase that provides α-complementation for blue/white color selection of transformants. The selection was based on the disruption of the *lacZ* gene by the insert. In the presence of the substrate X-gal (5-bromo-4-chloro-3-indolyl-β-D-galactopyranoside), transformants carrying the insert appeared white (see part VI for exceptions) since the X-gal was not metabolized. Those transformants that did not contain insert expressed β-galactosidase and turned blue due to the release of the blue indole dye. Colonies containing inserts were then used for further analysis as described in part VIII, Materials and Methods.

The cloned bands were then sequenced. DNA sequencing was performed using the cycle sequencing kit from Epicentre Technologies, Madison, WI. In cycle sequencing, a thermostable DNA polymerase is used for multiple rounds of DNA synthesis at high temperature. A set of four reactions generates sequence data from one strand (forward or reverse) of the template DNA. Each reaction contains a base-specific, chain terminating, dideoxynucleotide (i.e. ddATP, ddCTP, ddGTP, ddTTP), deoxynucleotides and [α-

^{32}P]dATP (10mCi/ml). The primers used for sequencing were M13, forward and reverse primers. Like T3 and T7 oprimers, M13 primers flank the multiple cloning site and encompass the insert. The polymerase extends the chain until a dideoxy nucleotide is incorporated which results in chain termination. The resulting radiolabeled nucleotides are then loaded in adjacent lanes of a polyacrylamide sequencing gel. Both forward and reverse strands were sequenced for Figure 2, part VI. The protocol for the sequencing reaction and sequencing gel is given in part VIII.

The inserts in the cloned bands were labeled as described in part VIII. The labeled probes were then used for Southern hybridization. In Southern hybridization, DNA fragments are denatured immobilized on a solid support medium such as nitrocellulose or nylon. The process of immobilization is referred to as "vacuum blotting". The probe is labeled, denatured to make it single stranded and hybridized to the membrane for several hours at a particular stringency. Stringency is a function of both, the ionic strength of the medium, and the temperature of hybridization. The blot is then washed in order to remove background hybridization, dried, and exposed to X-ray film. Each of these steps is outlined in part VIII.

III: RESULTS

Band purification for cloning

The polymorphic band of interest was cut out of the gel and reamplified according to the protocol described in Materials and Methods (Appendix). DAF 233 required only two rounds of re-amplification at higher annealing temperature (50°C) to obtain a pure band (Fig. 1A). The increased annealing temperature was used in order to increase the specificity of the reaction and eliminate any mismatched amplification products. With DAF220 however, a single band could not be obtained even after 5 rounds of re-amplification (Fig. 1B). The annealing temperature employed for DAF220 was 30°C since a higher annealing temperature reduced the yield drastically and did not significantly improve the purity (data not shown). This already suggested some oddity. Further purification of this band was carried out using a non-denaturing preparative polyacrylamide gel (Fig. 1B, lane NDG). This step yielded sufficient quantity of the fragment for cloning with elimination of bands higher than 140 bp, however, two smaller sized bands persisted. The integrity of the blunt-ended fragments after T4 DNA polymerase treatment (see Materials and Methods) was confirmed by running an aliquot on a DAF gel (Fig. 1A and B, lane DP).

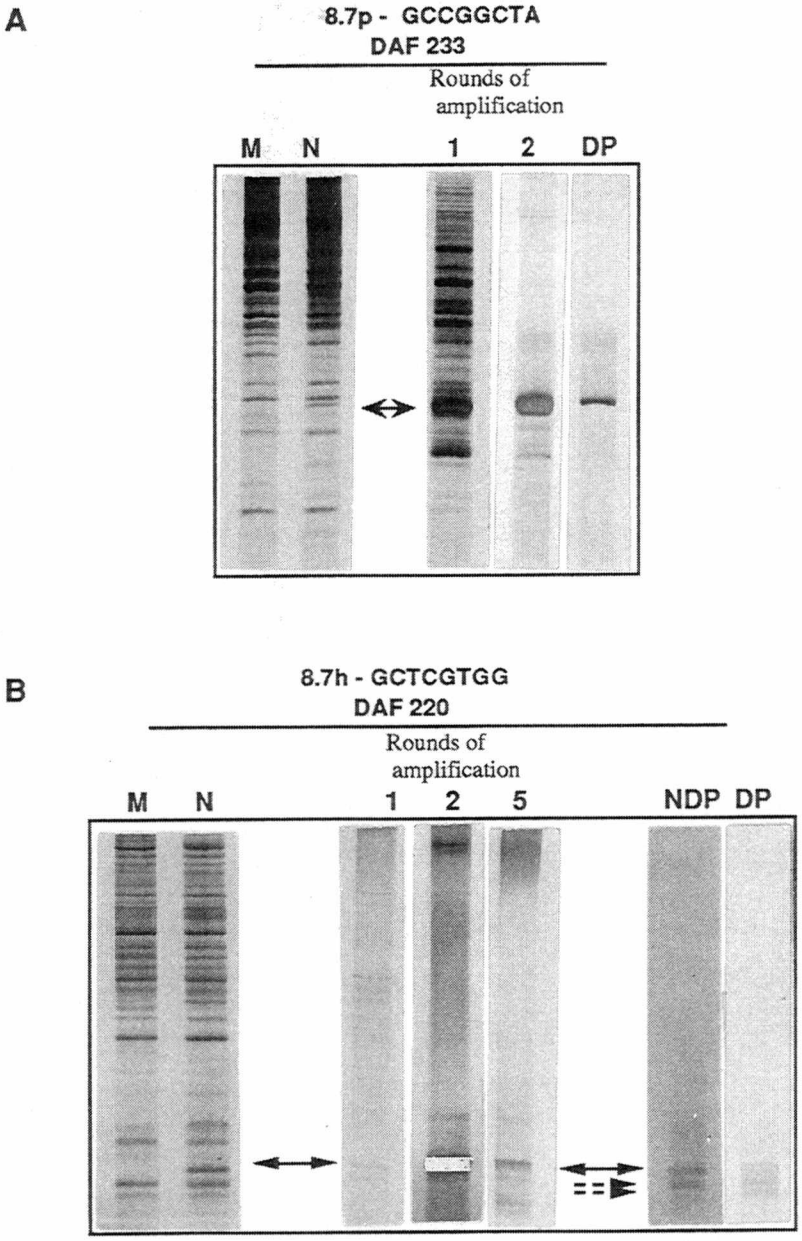


Figure 1: Purification of DAF233 and DAF220 from silver-stained polyacrylamide gels.
DAF233, the 185 bp polymorphic fragment (arrow) between 'Minsoy' (M) and 'NoirI' (N) using primer 8.7p, was re-amplified for two rounds (Panel A, lanes 1, 2).
DAF220, the 140 bp polymorphic fragment (arrow) between 'Minsoy' (M) and 'NoirI' (N) using primer 8.7h was reamplified for five rounds (Panel B, lanes 1, 2 and 5).
Subsequent purification on a non-denaturing polyacrylamide gel (NDP) resulted in co-purification of the fragment with smaller sized bands (dashed arrows).
Lane DP shows the fragment/s used for ligation after treatment with T4 DNA polymerase.
See Materials and Methods for details.

Strategies for cloning DAF products

To optimize yield, two different strategies were used to clone the DAF220 and DAF233amplification bands into the pBS^+ vector. A T-tailed method assuming that the extra 3' A residue was present in the fragment to be cloned and a blunt end cloning procedure. Table 1 shows the efficiency of transformation using the T-tailed and blunt-ended ligation, and the effect of phosphorylating the 5' end of the blunt-ended insert. Blunt end ligation gave a near 100 fold increase of colony forming units (cfu) or transformants compared to the T-tailed method. for example, transformation efficiencies for DAF233 were $0.16x10^4$ cfu/μg DNA (number 1) compared with $0.2x10^2$ (number 2). Blunt-end ligation gave comparable efficiencies for both DAF220 (number 3, $0.45x10^4$ cfu/μg) and DAF233 (number 1, $0.16x10^4$ cfu/μg). Phosphorylating the 5' end of the primer with kinase ($0.45x10^4$, number 3) resulted in a four fold increase over non-phosphorylation ($0.1x10^4$, number 4) for DAF220 with blunt end ligation. However, this increase was not always observed as in the case of DAF233 (number 1).

The efficiency of T-tailing was judged by the efficiency of self-ligation of the T-tailed vector compared to that of the *Sma*I cut precursor. Control T-tailed vector gave efficiencies of $0.3x10^3$ cfu/μg (number 6) whereas the efficiency of *Sma*I cut vector was $0.58x10^4$ cfu/μg (number 7). Self-ligation of the *Sma*I cut vector gave a 20 fold higher efficiency compared to the self-ligated T-tailed vector under similar conditions of cloning and transformation. Since there is no known restriction enzyme that produces a 'A' overhang, a qualitative estimation of the efficiency of T-tailing was considered the best available control. Also since the objective was to develop a quick and relatively easy cloning protocol that could be used for all DAF products, the use of other DAF fragments that contained a 3' A for control transformation was not considered. Control transformation with no DNA produced no transformants (Table 1, number 8). Positive control transformations with supercoiled plasmid gave consistent efficiencies between 0.4 and $0.6x10^5$ cfu/μg (number 5) between experiments indicating that the transformation protocol (see Materials and Methods) is reliable and that any differences in the efficiency between the two methods cannot be attributed to variable transformation efficiencies.

The presence of the correct insert was checked by using PCR directly on lac^- colonies. A random testing of 45 colonies of DAF233 and DAF220 (numbers 1 and 3, Table 1) showed that 90 - 100% of the colonies contained inserts. Out of those colonies containing

Table 1: Efficiency of cloning of DAF markers using two different methods

Ligation #	Insert	Ligation	Kinase treatment	Efficiency # cfu[3] per µg DNA	Lac⁻ colonies containing	
					insert[4] %	% correct insert
1.	DAF233	Blunt end	yes	$0.16x10^4$	90	80
2.	DAF233	T-tailed	yes	$0.2x10^2$	0	0
3.	DAF220	Blunt end	yes	$0.45x10^4$	100	70
4.	DAF220	Blunt end	no	$0.10x10^4$	50	75
5.	Control pBS[1]	Blunt end	-	$0.63x10^5$	-	-
6.	Control pBS[2]	T-tailed	-	$0.3x10^3$	-	-
7.	*SmaI* cut pBS	Blunt end	-	$0.58x10^4$	-	-
8.	Control transformation with no DNA	-	-	0	-	-

[1] Control pBS is supercoiled plasmid; efficiency for replicate was 0.4×10^5 cfu/µg
[2] Control pBS is T-tailed vector
[3] cfu colony transforming units (transformants)
[4] expressed as percent of total number of colonies

inserts, 70 - 80% contained the insert of predicted size (Table 1) whereas the remaining 20-30% consisted of multimeric insert fragments (data not shown). Twenty-five percent of the colonies containing insert were light blue on the primary agar plate and remained so upon replating (data not shown). Those colonies that showed a Lac$^-$ (white) phenotype on the primary plate but did not contain insert (10%, Table 1) turned blue upon replating.

Sequence analysis of cloned DAF220 and DAF233

Clones that showed an insert corresponding to the expected size were selected for sequencing (Fig. 2). The 187 bp DAF233 clone (Fig. 2A) showed complete primer sequences at the 5' ends. In case of DAF220, however, one clone showed the correct primer sequence at both ends (Fig. 2B, DAF220-A), whereas the other clone did not show the expected primer sequences at either end (Fig. 2C, DAF220-B). Partial primer sequences were found for DAF220-B at both ends; 4 out of 8 nucleotides at the 5' end and 3 out of 8 nucleotides at the 3' end. Sequence analysis using GCG software of all three clones DAF233 DAF220-A, and DAF220-B showed GC content of 29%, 51%, and 38% respectively (Fig. 2). A high GC content is an indicator of coding sequences. DAF220-A and DAF220-B did not show any similarity to each other in terms of their sequence or GC content. The sequences did not show any identity to vector sequences and did not contain any tandem repetitive motifs found in plants [Wang et al., 1994]. Comparison of the three sequences to those in the GenEMBL showed no identity to soybean sequences contained in the gene bank. DAF233 showed a 60% identity in a 120 bp overlap to *Marchantia polymorpha* (liverwort) choroplast genome, DAF220-A showed a 75% identity in a 32 bp overlap to the large subunit of RNA polymerase II from *Arabidopsis thaliana* and DAF220-B showed a 74% identity in a 27 bp overlap to rice chloroplast genome. Analysis of all possible open-reading frames indicated that only DAF220-A contained an open reading frame that encoded 45 amino acids. The absence of a stop codon at the end of the DAF220-A suggests that the sequence could be a part of a larger coding sequence.

Genomic analysis of DAF amplicons

In order to determine if the mapped DAF markers originated from single copy or repetitive sequences in the genome, the cloned fragments (DAF233 and DAF220-A) were hybridized to a blot containing genomic DNA from Minsoy and 'Noir1' that had been restricted with different enzymes. Results from Southern hybridizations to genomic DNA (Fig. 3A, BLOT-GEN) obtained with DAF233 showed monomorphic fragments with *Hind*III (5 kb)

A) DAF233

```
                10                        30                        50
  5' GCCGGCTATA GGTAGTAATT AAGATGACCT AGATTTTATC ATCGACATAA

                          70                        90
     TATGTGATAA TAAAAAAAAA TTATTTGGAT GAATTTCCAT CATTTTCTGT

               110                       130                       150
     ATTTTGTTGT CCTTGTTAAT TTCTTTTTCG TCTTGTAACT TTTTAATGGG

                          170                  187
     AAGAAATGAA ATATCCATAT TAAATCACCT AGCCGGC  3'
```

A=73*; C=29; G=26; T=59; A+T= 70.6%; G+C=29.4%; AAA=15; TTT=10; CCC=1; $(A)_n=A_9$

B) DAF220-A

```
                10                        30                        50
  5' GCTCGTGGCT TACGATCGAT TTTCCCGCGA TTTCTTTTCG CCTTCCAACG

                          70                        90
     TCGATCCGTC GTTCGTTCCG TCGAAAACGT CGTAGCCCAA GAAACTTAAA

               110                       136
     TTTGCTCCCT GAGCGAAGAC AATATGGCCC ACGAGC   3'
```

A=37; C=29; G=41; T=29; A+T=48.5%; G+C=51.5%; AAA=6; TTT=4; GGG=4

C) DAF220-B

```
                10                        30                        50
  5' GTGGTGCTCG TAATGATGGG ATTCCATGAT TATGGAGAGG TTGATAAGGT

                          70                        90
     ATATGCCAAA AAAATTATCC CATTAATGTC ATTTAGGATA ATTCTCTCGT

               110                       135
     GTGTAACGAC ATTACATACC ATTGCAGTGG ATCCA    3'
```

A=43; C=30; G=21; T=41; A+T=62.2%; G+C=37.8%; AAA=1; TTT=1; GGG=1; CCC=1

Figure 2: Sequences of cloned polymorphic fragments DAF233, DAF220-A, and DAF220-B. Underlined sequences indicate whole or partial primer sequences. Expected primer sequences for DAF233 and DAF220 (A and B) are GCC GGC TA and GCT CGT GG respectively.

* refers to the number of bases except when expressed as a percentage

and *Taq*I (0.6 kb). When hybridized to a blot of DAF profiles from Minsoy and Noir 1, DAF233 showed a single polymorphic band present in Noir 1 corresponding to 185 bp (Fig. 3A, BLOT-DAF). In case of DAF220-A, (Fig. 3B, BLOT-GEN) Southern hybridization to genomic DNA showed bands with *Eco*RI (3.5 kb), *Eco*RV (2.7 kb) and *Taq*I (not shown) and hybridization of DAF220-A to a DAF blot of profiles from Minsoy and Noir 1 showed a single band at 140 bp in both parents (Fig. 3B, DAF-BLOT).

IV: DISCUSSION

Polymorphic markers from silver-stained gels are purified from DAF profiles that contain up to 100 amplification products. The band purification method of fragment excision and re-amplifying is associated with varying degree of success depending on the type of DAF product being purified. DAF233 was successfully purified in two rounds of amplification but DAF220 could not be purified even after 5 rounds of purification on 6% polyacrylamide gels. These extra bands in DAF220 arise either from amplification artifacts that are extremely reproducible, or from contaminating bands present in the desired band that are efficiently amplified, or from internal primer sites in the DAF product. However, sequence analysis showed no internal primer sequences in DAF233 or DAF220-A.

Two different variants of *Taq* polymerase were used to amplify the two DAF markers; the truncated Stoffel fragment, for DAF233 and Ampli*Taq*® for DAF 220 (see Materials and Methods in part VIII). It is possible that Ampli*Taq*® is more sensitive to small quantities of contaminating templates compared to Stoffel fragment and this may account for the differences observed in case of DAF220 and DAF233 purification. The source of additional bands for DAF220 is not from the enzyme since control amplification reactions with no added DNA showed no products with both DAF220 and 233 (data not shown).

Blunt end ligation gave a 100 fold increase in the number of colony forming units compared to the T-tailed method. This result is consistent with recent findings suggesting that the extra nucleotide added by *Taq* polymerase need not exclusively be an A residue as generally assumed [Costa and Weiner 1994]. The nucleotide added at the 3' end depends on the type of nucleotide present at the 5' end of the primer. If the 5' nucleotide is a G, C, or A, the extra nucleotide can be A, C, G, or T. Only if a 5' T is present in the primer, an extra A residue is transferred by the polymerase [Costa and Weiner 1994]. The primer sequence that generated DAF233 has a 3' C at the 3' end of its primer sequence and could have an extra G, A, or C at the 3' end. This may explain the poor efficiency of ligation with

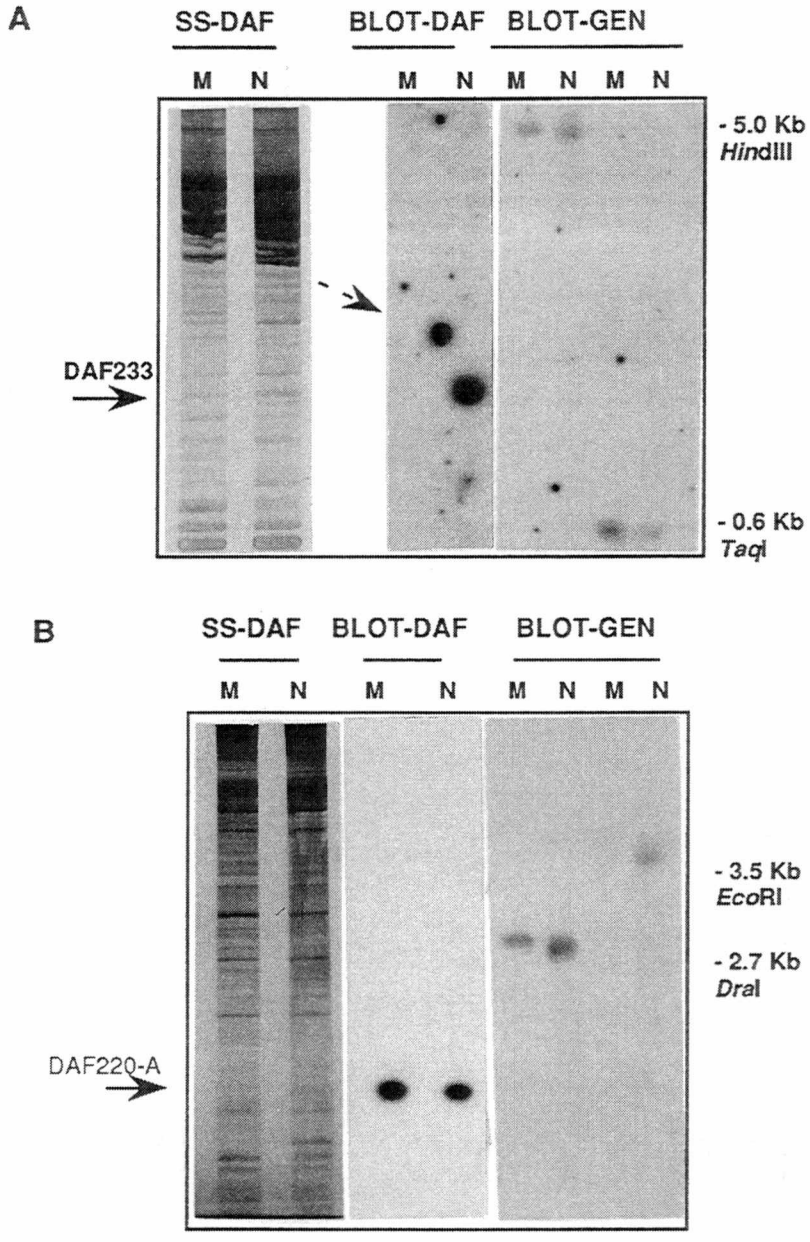


Figure 3: Southern hybridization of cloned DAF markers.
Cloned DAF markers, DAF233 (Panel A) and DAF220-A (Panel B) that were polymorphic between Minsoy (M) and Noir 1 (N) on silver stained gels (SS-DAF) were labeled and hybridized to a blot containing DAF profiles (DAF-BLOT) of Minsoy and Noir 1 and to genomic blots (BLOT-GEN) of Minsoy and Noir 1. DAF233 hybridized to a 185 bp DAF product from Noir 1 on the DAF blot , a 5 kb *Hin*dIII fragment and a 0.6 kb *Eco*RI fragment on the genomic blot. DAF220-A hybridized to a 140 bp DAF product from Minsoy and Noir 1 on the DAF blot, a 3.5 kb *Eco*RI fragment and a 2.7 kb *Dra*I fragment on the genomic blot. The dashed arrow on the DAF blot in panel A is background signal.

the T-tailed vector. Phosphorylation of the 5' end of the primer may increase the efficiency of blunt-end ligation. Although this represents an extra step in the procedure it takes only 15 minutes, is performed in ligase buffer, no phenol-chloroform extractions are involved and it eliminates the need to purchase more expensive 5' phosphorylated arbitrary primers. However, replicate experiments are necessary to determine if the observed increase in cloning efficiency is significant.

The efficiency of blunt-end cloning described here ($0.16x10^4$-$0.45x10^4$ cfu/μg) is at least three fold higher compared to a published protocol ($0.05x10^4$) by Liu and Schwartz [1992], although the two protocols are comparable as regards the ease and reagents required. The cloning protocol presented here, however, is 100 fold lower in efficiency compared to a protocol that utilized linker and adaptors (10^6 cfu/μg) [Lorens et al., 1991] and is approximately ten fold lower than published methods of T-tailing ($0.24x10^5$) [Kovalic et al., 1991]. These published methods, however, involve many steps and require the use of special vectors. Colonies containing inserts of the predicted size were white or light blue which could occur due to the small size of the inserts (187 and 136 bp). Since the *Sma*I site is located at the beginning of the B-galactosidase coding sequence in the pBS^+ cloning vector, it is possible that the extra 50 amino acids allow for a tertiary structure of the enzyme that retains a reduced enzymatic activity resulting in a leaky Lac^- phenotype. When dealing with small sized inserts therefore, screening light blue colonies is recommended.

Sequencing revealed that some cloned DAF markers contain truncated primer sequences at the ends of the fragment. The partial primer sequences did not arise from a loss of bases during "polishing" with T4 DNA polymerase since a decrease of 7 bases would have produced a change in the molecular size which would be detected on a mini-gel. No such reduction in size was detected. A possible explanation for this observation is that these irregularities could arise during cloning at the junction of the vector and insert with particular primer sequences only. The lack of identity of DAF233, DAF220A, and DAF220-B with soybean DNA is not surprising since the total number of soybean sequence data (in bp) in the GenEMBL databank does not exceed 200 kb [Wang et al., 1994]. In case of DAF220-A, the absence of a high AT content, the presence of an open reading frame, and identity to the large subunit of RNA polymerase suggests that DAF220-A has the potential to be a marker of significance, despite a small 15 bp identity. The lack of an open reading frame in DAF233 and DAF220-B, and the high AT content of the

sequences indicates that these markers originate from non-coding DNA. In addition, hybridization of DAF220-A to low copy DNA suggests that DAF markers therefore can originate from coding sequences.

Hybridization of DAF233 to the DAF blot confirmed that the correct polymorphic fragment was cloned and indicates that DAF233 is a unique amplified product that is not codominant to any other band present in the profile. The banding patterns produced by DAF233 on genomic blots did not resemble patterns produced by repeated DNA and suggests that DAF233 originated from low copy nuclear DNA consistent with the sequence analysis showing a lack of repeated sequences.

A Southern blot of DAF 220-A was also consistent with the origin of DAF220-A from low copy DNA. When hybridized to a blot containing DAF products, DAF220-A revealed monomorphic bands, i.e., bands of equal intensity and identical molecular weight were found in both parents. This indicates that (1) there is a smaller cross hybridizing monomorphic band present on an agarose gel but not on an acrylamide gel which is consistent with no polymorphic bands seen on an agarose gel (data not shown), or (2) that the sequence corresponding to DAF220-A must be present in Minsoy (which shows the null allele of DAF220). The latter explanation can be eliminated based on control experiments in which the gel piece containing no band from Minsoy was amplified but no band corresponding to the size of DAF220 was detected (data not shown). Since the cloned polymorphic fragment DAF220, produced two different clones (DAF220-A and B) that did not show any sequence identity, suggests that the polymorphic fragment DAF220 contains sequences of multiple origin. The proportion of DAF220-A and DAF220-B will have to be determined by analyzing and sequencing several different clones. The crucial experiment showing a polymorphic band when hybridizing a clone to a DAF blot will identify the true polymorphism between Minsoy and Noir 1. If DAF220-B shows a polymorphism on a Southern blot of DAF products, it would indicate either that the polymorphism is due to an altered primer site or that it is due to an insertion. This would mean that the co-dominant DAF product larger than 2-3 kb and is not efficiently amplified (a deletion seems unlikely as the original polymorphism is only 140 bp). Furthermore, a large insertion could produce a restriction fragment length polymorphism on a genomic Southern.The sequence of DAF233 can be used to design primers for the purpose of converting it into a "locus specific" SCAR. A lack of polymorphism with the SCAR primer can be converted into a "true polymorphism" by restricting the PCR fragment with four base cutters [Paran and Michelmore 1993, Caetano-Anollés et al., 1993].

PART VII

SUMMARY AND FUTURE ASPECTS

SUMMARY AND FUTURE ASPECTS

My work adds to the growing body of evidence that DAF is a valid technique for mapping, fingerprinting, estimation of relatedness of germplasm, and generating SCARs. The utility of DAF markers for mapping was demonstrated in this study by showing Mendelian inheritance in inter- and intra-specific crosses of soybean. Three DAF markers were mapped in the soybean intraspecific (recombinant inbred) map of 'Minsoy' and 'Noir1'. One marker mapped to a QTL region associated with seed pod filling. Two markers revealed by the same primer mapped to different sites in the genome confirming that DAF products originate from arbitrary sites in the genome. In the interspecific *G. soja* x *G. max* cross, approximately 25% of DAF markers showed non-Mendelian inheritance. One possible organelle marker and possibility of paternal inheritance of another marker was shown. The origin of these markers needs confirmation by cloning and hybridization to genomic and organelle DNA. The hybridization of DAF233 confirmed its polymorphic nature and can now be converted into a SCAR by using its sequence to design SCAR primers. Linking different maps of a species will be an important focus for map utilization, and DAF markers can play an important role in this area by providing numerous polymorphisms in all regions of the genome.

The genetic basis of polymorphisms is not fully understood and no data exist on this aspect. Results with DAF220 showed that polymorphic bands can contain sequences of multiple origin. Conversion of SCARs will not be straightforward in such cases and it serves as a cautionary note for conversion of DAF markers into SCARs, especially for using DAF markers for transferring to other crosses. A very small percentage of APT primers have been shown to be codominant markers. Cloning and sequencing of genomic clones from which the polymorphism originated, can provide answers to primer site specificities and may reveal the codominant nature of the polymorphism in the genome. Genomic clone characterization will also resolve the question whether template-primer mismatches produce variation in intensity of bands in DAF profiles. If factors influencing the outcome of an amplification process are better understood with information from genomic clones, this extends our knowledge for design of "locus specific" primers for mapping applications.

APT markers have been stigmatized by several shortcomings of the technology like reliability, repeatability, dominance, low allele number (only 2), lack of homology among

related taxa and lack of specificity for unique regions. These issues have been addressed in this thesis and DAF was shown to overcome some of them. DAF is reliable, reproducible, has a high discriminatory power and generates a high level of information. Both cloned DAF bands were shown to originate from low copy sequences in the genome suggesting that DAF markers do not show particular bias for repeated regions of the genome. Information on the proportion of DAF markers originating from repeated sequences will be forthcoming from the growing database of characterized DAF markers. DAF has a limitless supply of markers defined only by the combination of arbitrary sequences for primers. Much fewer primers are needed for screening compared to RAPD primers resulting in tremendous savings in time and resources.

Management of germplasm has scientific, legal and practical implications, and the technique of DAF will likely be an important methodology in all three areas. DAF can distinguish agronomically important cultivars and was as sensitive as RFLP for estimating genetic relationships with a small number of primers. The choice of DAF for determining genetic relatedness among genotypes will depend on studies in several species showing that these markers are capable of providing comparable results compared to other techniques like RFLP, allozyme or morphological markers. Results presented in this thesis and those with turfgrass are promising in this regard.

DAF or APT markers in general will find a niche in two main areas of plant genome biology, map-based cloning and fingerprinting. Aspects of map-based cloning extends to finding linked markers to important genes using BSA or NILs, or generating "high volume" markers (SCAR or STS) for saturating a region for positional cloning. Aspects of fingerprinting technology includes areas of varietal and genotype identification especially if automation is available. Automation will also make "marker assisted breeding" a much more accessible and viable alternative.

Besides map-based cloning of the *nts* gene, techniques like differential display and RT-PCR (Reverse Transcriptase-PCR) are valuable techniques to show differences between *nts* and Bragg. The strategy of finding a protein difference between *nts* and Bragg can be extended to finding cDNA differences with these techniques and may expedite the search for the *nts* gene. The work presented in this thesis demonstrates the validity of DAF to reveal molecular markers that are suitable alternative markers for saturating the *nts* region and identifying YAC clones.

PART VIII

MATERIALS AND METHODS

I: TWO-DIMENSIONAL PROTEIN GEL ELECTROPHORESIS

Plant materials and growth conditions

Plants were either grown in plastic pouches or pots containing vermiculite or a mixture of sand-vermiculite. Non-sterilized vermiculite or sand-vermiculite-based growth conditions produced occasional nodulation of uninoculated plants due to contamination. Sterilizing the sand-vermiculite for 10 min at 120°C efficiently eliminated *Bradyrhizobium* but longer sterilization periods caused Mn toxicity in the plants. Cultivar Bragg has especially high manganese toxicity (D. Eskew, pers. comm.).

Growth of the plants in plastic pouches was best for the first 25 days of growth. Leaf material was never harvested from pouch-grown plants older than 21 days after inoculation. The seeds were surface sterilized with hypochlorite and thoroughly rinsed with sterile water. For pot-grown plants, the seeds were sown directly into the sterile medium and placed in a greenhouse under controlled conditions. Light cycles were 16 h light and 8 h dark. Four to seven days after germination, the plants were inoculated with *Bradyrhizobium japonicum* strain USDA110 (0.5 ml per pouch-grown plant at 10^8 cells/ml or diluting 50 ml of the same culture to 4 liters with nutrient BMM solution and added to the pot-grown plants).

Growth of Bradyrhizobium japonicum USDA110

Bradyrhizobium japonicum USDA110 was grown in BMM medium at 28-30°C for five days until optical density at 600 nm was equal to one Klett unit. This culture was diluted 1:10 before inoculation. In mg/L the medium, BMM, contained; $Na_2HPO_4.12H_2O$, 360; $MgSO_4.7H_2O$, 80; $FeCl_3.6H_2O$, 3; $CaCl_2.2H_2O$ 40; trace elements, $MnSO4.4H_2O$, 0.1; H_3BO_3, 0.03; $ZnSO_4.7H_2O$, 0.03; $NaMoO_4.2H_2O$, 0.0025; $CuSO_4.5H_2O$, 0.0025; $CoCl_2.6H_2O$, 0.0025; biotin, 0.1; thiamine, 1; mannitol, 10,000; sodium glutamate, 500; Difco Bacto Agar, 10,000; pH 6.8-7.0.

Total in vivo protein extraction

Total *in vivo* proteins were extracted according to Flensgrud and Kobro [1989] with a few modifications. Leaf tissue (0.5 to 0.7g) was mixed with an equal weight of polyvinyl-pyrrolidone and ground in liquid nitrogen. Nine ml of extraction buffer (50 mM pyridine, 10 mM thiourea, 1% SDS, 1 mM Na_2EDTA and 2 mM phenylmethylsulfonyl fluoride, pH

5) was added to the mixture. The sample was centrifuged at 16,000 rpm (Beckman, JA20 rotor) for 40 minutes at 4°C. Acetone at -20°C was added to the supernatant to a final concentration of 90% and the supernatant was stored at -20°C overnight. The precipitate was collected by centrifugation and dried in a stream of nitrogen. The precipitate was dissolved in 500 µl of buffer (9.0 M urea, 1.6% ampholytes pH 3-10, 0.4% ampholytes pH 5-7, 2.5% dithiothreitol). The samples were dialyzed for 2 hours against 25 ml of the same buffer containing 2% lysine instead of ampholytes. To the sample, 2% ampholytes (as above) and CHAPS (3- [(3-cholamidopropyl)dimethylammonio]-1-propanesulfonate) was added to a final concentration of 26:1 (w/w) and analyzed immediately by 2-D gel electrophoresis or stored at -70°C for later use. Based on estimates of total protein content in soybean leaves [ ], the efficiency of the extraction protocol was 80%.

In vivo labeling of total plant with ^{35}S

Bragg seeds were surface sterilized with 10% chlorox for 2 minutes and thoroughly rinsed with sterile deionized water. Plants were grown in styrofoam cups containing autoclaved sand in a growth chamber with a 16h/25°C : 8h/20°C, day : night cycle. The plants were watered with 100 ml Herridge solution at the time of planting, 50 - 100 ml sterile distilled water every alternate day and modified Herridge containing one tenth concentration of sulfate (50 mM) for labeling with ^{35}S ($^{35}SO_4$ was from DuPont NEN, MA). Seven day old plants were inoculated with a 4 day old culture of *Bradyrhizobium japonicum* USDA110 (10^9 cells per pot). 1 mCi of ^{35}S in modified Herridge was added to each plant at different times after inoculation. After 24 hours of labeling, the plant was removal from the sand, and kept in ice water before harvesting the various tissues. The tissues were weighed separately and immersed in liquid nitrogen as quickly as possible. The labeled plant tissues were stored at -70°C until further use.

A 1000X solution of Herridge solution contained per liter, KH_2PO_4, 17 g; K_2HPO_4, 21.8 g; KCl, 18.7 g; $MgSO_4.7H_2O$, 123.3 g; $CaCl_2.2H_2O$ 36.7 g; Ferric monododium salt of EDTA, 8.7 g; H_3BO_3, 0.715 g; $MnCl_2.4H_2O$, 0.453 g; $ZnCl_2$, 0.028 g; $CuCl_2.2H_2O$, 0.013 g; $NaMoO_4.2H_2O$, 0.006 g; $CoCl_2.6H2O$ 0.01 g; Ni, 0.1.16 ml of 1000 ppm (1000 ppm is 17 mM); KNO3, 0.05 g.

Two-dimensional gel electrophoresis

Protein products were analyzed by 2-D gel electrophoresis on a Bio-Rad Protean or mini-Protean gel apparatus according to the procedure of Hochstrasser et al., [1988]. The

concentration of total proteins extracted from Bragg and nts were measured using the Bio-Rad assay and equal amounts (200 μg) were loaded in the first dimension. For the *in vivo* labeled proteins, approximately 5 - 9 x 10^5 cpm of a TCA precipitate containing 50 μg was loaded in the first dimension. A pH gradient from 3 to 10 was produced by a mixture of ampholytes (1.2% ampholytes pH 5-7, 0.4% ampholytes pH 3 -10 and 0.4% ampholytes pH 4-6; Pharmacia Biotech, Uppsala, Sweden). The establishment of the pH gradient was verified by measuring the pH of different sections of a control tube gel according to Lan and Chrambach [1981]. A 12.5% SDS-polyacrylamide slab gel was used in the second dimension (^{14}C molecular weight standards between 14 and 200 kD were obtained from Bethesda Research Laboratories, NY). The dried *in vivo* gels were exposed to XOMAT-AR X-ray film at -70°C for 7 to 15 days and the film developed according to the directions of the manufacturer (Kodak).

II: GENOTYPES USED FOR STUDYING INHERITANCE AND MAPPING OF DAF MARKERS

The genus *Glycine* subgenus soja is made up of two species; *G. soja* Sieb and Zucc, the wild progenitor and *G. max*, the cultivated species. Both subspecies have a chromosome number of 2n = 40 and have been shown to hybridize readily with each other generating vigorous and fertile F1 plants. Pachytene analysis revealed that *G. soja* and *G. max* carry similar genomes [Singh and Hymowitz 1988].*G. soja* refers to *Glycine soja* PI468.397 and *G. max* refers to *G. max* line nts382 [Carroll et al., 1985b] except where indicated. *G. max* cv Bragg is referred to as Bragg.

The supernodulating mutant *G. max* line nts382 was produced by EMS-mutagenesis of *G. max* (L.) Merr. cv Bragg [Carroll et al., 1985a]. Compared to the wild-type Bragg, the supernodulating mutant soybean line nodulates upto 40 times more and produces nodules in the presence of otherwise inhibitory levels of nitrate [Carroll et al., 1985a,b]. The *nts* mutation is recessive, and segregates as a single Mendelian locus [Landau-Ellis et al., 1991]. The generation of the original cross, *G. max* line nts382 x *G. soja*, F1 and F2 progeny are described in [Landau-Ellis et al., 1991]. F1 hybrids were confirmed by RFLP banding pattern and a phenotype that was intermediate between *G. soja* and *G. max* nts382 [Landau-Ellis et al., 1991]. From a total of 113 F2 plants, 24 segregated for the mutant phenotype confirming a 3:1 (wild-type to supernodulating) Mendelian inheritance of the *nts* gene. RFLP markers also segregated 1:2:1 (D. Landau-Ellis and P. Gresshoff, unpublished data) confirming the utility of the population. For this DAF analysis, a random subsample of 36 plants was chosen. This subsample segregated at 24 wt : 12 nts (27:9

expected). For further mapping purposes, an immortal population (F11) of recombinant inbred lines (RIL) resulting from a cross of *Glycine max* (L.) Merr. cv Minsoy (PI 27.890) and *Glycine max* (L.) Merr. cv Noir 1 (PI 290.136) was used [Lark et al., 1993].

III: DNA AMPLIFICATION FINGERPRINTING

The technique of DAF involves DNA isolation, DNA amplification, and separating the products on a silver stained polyacrylamide gel. Each of these procedures is described below.

DNA isolation and quantification

Genomic DNA was isolated from young soybean leaves according to Dellaporta et al., [1983]. Young soybean leaves (0.5 g) were ground in liquid nitrogen and 0.3 to 0.5 g Polyclar AT (Polyvinylpyrrolidone) from SERVA was added to the powdered leaf and liquid nitrogen mixture. The cold powdered leaf mixture was put into a 50 ml plastic centrifuge tube and 15 ml of extraction buffer (0.1 M Tris HCl pH 8.0, 50 mM EDTA, 500 mM NaCl) was added to it. One ml of 20% SDS was added, the mixture was gently mixed and incubated in a 60°C water bath for 1 hour. After incubation, 5 ml of 5 M potassium acetate (pH 8.0) was added and the tubes were allowed to sit on ice for 30 minutes. The mixture was filtered through a fluted filter paper into clean plastic tubes. Ten ml of cold isopropanol was added to the filtrate and left at -20°C for half an hour. The solution was centrifuged for 15 minutes at 4°C using a JA20 rotor at 11,000 rpm. The supernatant was poured off and 0.7 ml chromosomal DNA buffer (50 mM Tris-HCl ph 8.0, 10 mM EDTA) was added to the precipitated DNA. The DNA was put on ice to dissolve (if not dissolved then warmed at 60°C for a few minutes in order to dissolve). The DNA solution was put into an eppendorf tube and centrifuged for 10 minutes at 12,000 rpm to precipitate out polysaccharides and SDS complexes. The supernatant was put into a new eppendorf tube and 3 M sodium acetate (pH 7.5) at one-tenth the supernatant volume (75 µl) was added. Isopropyl alcohol (500 µl) was added to the tube and mixed gently. The tube was spun at 14,000 rpm (maximum speed) in an eppendorf centrifuge for two minutes. The supernatant was poured off and the DNA was washed with 70% ethanol (500 µl) to remove salts. The tubes were kept at room tempaerture for 5 minutes and spun again at maximum speed for two minutes. The ethanol was poured off and the pellet was dried in a speed vacuum concentrator. The pellet was dissolved in 100 µl sterile distilled water and stored at 4°C.

DNA concentration was measured in a TKO100 Dedicated Mini-Fluorometer (Hoeffer

Scientific Instruments, San Fransisco, CA). The assay is based on fluorescent enhancement of the dye, H33258, which is highly specific for double stranded DNA [Hoeffer instruction manual]. Since the dye has reduced affinity for single stranded DNA or RNA, double stranded genomic DNA can be accurately quantified.

DNA Amplification Fingerprinting

Amplification reaction was carried out in 25 µl volume containing a total of 5 ng of template DNA, 3 µM oligonucleotide primer (National Biosciences, Plymouth, MN and Integrated DNA Technologies), 1 unit of Taq DNA polymerase (AmpliTaq® Perkin-Elmer/Cetus), 200 µM of each dNTP, 10 mM Tris-HCl (pH 8.3), 50 mM KCl and 2.5 mM $MgCl_2$. The reaction mixture was overlaid with mineral oil (Mallincrodt, St. Louis, MO). Amplification was carried out in a thermocycler (Ericomp Inc., San Diego, CA) with 2 cycles of 95°C, 30°C , 72°C for 1 minute each, 33 cycles of 95°C, 36°C, 72°C for 20, 40, 20 seconds respectively, and 1 cycle at 72°C for 5 minutes. These amplification parameters give essentially identical results to the ones described by Caetano-Anollés et al., [1991b]. Amplifications with all primers used AmpliTaq® enzyme except for one set of experiments with primer 8.6a (Table 1, part 4) in which Stoffel fragment of Taq polymerase (Perkin-Elmer/Cetus) was used. Stoffel fragment is a recombinant DNA polymerase encoded by a modified form of the *Thermus aquaticus* DNA polymerase gene. Reaction and cycling conditions for Stoffel fragment are described in Caetano-Anollés et al., [1991d]. Amplified products (3 - 4 ml) were separated on 4.5% polyacrylamide gels at 110 V under constant voltage using 1X TBE buffer (0.1 M Tris base, 0.082 M boric acid, 1 mM EDTA) and detected by silver staining [Bassam et al., 1991].

For silver staining, the plastic backed gels were first fixed for 15 minutes with 7.5% acetic acid, washed three times with distilled water for 2 minutes each, impregnated with silver solution (1 g/L $AgNO_3$, 1.5 ml/L 37% formaldehyde) for 20 minutes, rinsed very briefly (5 seconds) with distilled water and developed using a solution (30 g/L Na2CO3, 3 ml/L 37% formaldehyde, 2 mg/L thiosulfate). Development was stopped by the addition of fixing solution (7.5% acetic acid) when the gels showed dark bands on a pale background).

DAF data analysis

The data presented in DAF gels were scored by visual observation. Standard errors for band mobility varied by an average of 0.2% for bands scored between 100 - 700 bp [Gresshoff and Mackenzie 1994]. Laser densitometry scans were performed on Ultroscan

XL (LKB Produkter AB, Bromma, Sweden). Scans were evaluated using Gelscan XL version 2.0 software (Pharmacia LKB Biotechnology, Uppsala, Sweden). The primers were coded to indicate length, %GC content and its numerical position in the particular set. For example, 8.6n indicates a primer that is 8 nucleotides long, containing 62.5% GC and occupies the n^{th} position among our stock of primers of the same length and GC content. In part III, for presentation purposes, entire lanes are not shown in Figures 3 and 4. In part III, for comparison purposes, lanes 1, 2 (Fig. 1) have been shown again in Fig. 2 (GS, GM; primer 8.10b) and Fig. 5, (*G. soja* and *G. max*).

IV: MAPPING DAF MARKERS

Minsoy, Noir 1 and recombinant inbred DNAs were supplied by Dr. G. K. Lark, University of Utah. The parental DNA, Minsoy and Noir 1 was used directly for amplification as described in the preceeding sections and screened for polymorphisms. The RI lines were then amplified to obtain segregation data of the DAF marker. Mapping data was sent to Dr. Lark for analysis since the data base is at the University of Utah. The segregation data were analyzed using the MAPMAKER program which uses two-point and three-point linkage analysis. Further description of the MAPMAKER program can be found in Lander et al., [1987]. LOD is an abbreviation for Log Likelihood Difference and is a measure of the probability of a particular order of loci. With reference to Table 2, part IV, a LOD score greater than three indicates linkage.

V: METHODS USED FOR GENOTYPE DISTINCTION AND PEDIGREE ASSESSMENT OF PIONEER HI-BRED SAMPLES

The genotypes used for pedigree assessment were of unknown relationship to the Knoxville researchers were supplied by Pionner Hi-bred, Johnston, IA. The seeds were planted in the Racheff greenhouse.Young trifoliate leaves were harvested from 14 day old plants and DNA (coded 1 to 10) was extracted according to Landau-Ellis et al., [1991].

Cluster analysis of the data was obtained at Pioneer Hi-Bred using the S-PLUS software that uses average linkage method and is used here with the permission of Drs. S. Smith and D. Webb. Pedigree and RFLP data were also supplied by Pioneer Hi-Bred. The information about the RFLP probes was not available at the time of writing this dissertation.

VI: CLONING DAF POLYMORPHIC BANDS

The reader is advised that some portions of the methodology for cloning, sequencing and Southern hybridization is repeated in this part for sake of clarity and continuity. For cloning DAF fragments, the polymorphic band was first purified from the gel, and then cloned into the pBluescript (pBS^+) plasmid. The pBS^+ vector along with the ampicillin resistance gene, contains the *lacZ* gene coding for β-galactosidase that provides α-complementation for blue/white color selection of transformants. The selection was based on the disruption of the *lacZ* gene by the insert. In the presence of the substrate X-gal (5-bromo-4-chloro-3-indolyl-β-D-galactopyranoside), transformants carrying the insert appeared white (see part VI for exceptions) since the X-gal was not metabolized. Those transformants that did not contain insert expressed β-galactosidase and turned blue due to the release of the blue indole dye. *Band amplification and purification*

In this study, parental soybean genotypes, *G. max* cv Minsoy and cv Noir 1, used to construct the genetic map developed at the University of Utah [Lark et al., 1993] were used. DAF 220 and DAF 233 have been mapped to linkage groups 4 and 10 [Prabhu and Gresshoff 1994] in this intraspecific map [Lark et al., 1993]. Isolation of DNA and amplification conditions are identical to those described in Prabhu and Gresshoff [1994] except for purification of marker DAF 233, wherein an annealing temperature of 50°C was used. Band purification was performed using 6% polyacrylamide gels as described in Weaver et al., [1994]. Non-denaturing polyacrylamide electrophoresis was carried out according to the protocol described in Maniatis without modifications [Sambrook 1989].

Manipulation and Cloning of Vector

All restriction enzymes used in this cloning procedure were purchased from New England Biolabs, Beverly, MA. pBS^+ was obtained from Stratagene, La Jolla, CA. The vector pBS^+ (3.2 kb) was purified on a cesium-chloride gradient as described in Maniatis [Sambrook 1989]. For *Sma*I restriction, 10 µg of vector was mixed with 2 µl of 10,000 units/ml *Sma*I, 2 µl 10X *Sma*I buffer in a total volume of 18 µl. The mixture was incubated at 28°C for 2 hours. The *Sma*I was not deactivated before use in the ligation reaction. The *Sma*I cut vector was kept at -20°C until further use.

For T-tailing, to 3.5 µg of *Sma*I cut plasmid was added, 10 µl 10X PCR buffer, 10 µl 25 mM $MgCl_2$, 1 µl 2 mM dTTP, 0.6 µl AmpliTaq® (5 units/µl) in a total volume of 100 µl. The mixture was incubated at 70°C for 2 hours and phenol-chloroform extracted before

ligation.

Pretreatment of Inserts

For blunt-end ligation, the insert was treated with T4 DNA polymerase in order to remove the extra adenosine residue. To 150 ng insert was added 0.5 µl of 2 mg/ml BSA, 1 µl 2 mM dNTPs, 0.33 µl T4 DNA polymerase (3000 units/ml) in a total volume of 22 µl. The mixture was incubated at 11°C for 20 minutes and 75°C for 10 minutes to inactivate the polymerase. The insert DNA was extracted with phenol-chloroform before use. To 100 ng of insert (140 bp) was added 20 units of T4 polynucleotide kinase and 0.8 µl 10X ligase buffer in a total volume of 10 µl. The mixture was incubated at 37°C for 15 minutes and heat inactivated for 20 minutes at 65°C.

Ligation and Transformation

To the insert described above was added, 50 ng vector (1 µl) 0.5 µl ligase (40,000 units/ml) and 0.2 µl 10X ligase buffer to make up the volume to 10 µl. The ligation mixture was incubated at 16°C for 16 hours. The ratio of pmol ends of insert to vector was kept at 50:1 for all ligations. Transformation was carried out according to Maniatis using the $CaCl_2$ method.

Analysis of Recombinant Clones

Lac^- colonies were used directly for PCR analysis using T3 and T7 primers which are complimentary to a region flanking the multiple cloning site and encompass the entire length of the insert. After streaking the colony on a $X\text{-}Gal^+$, Amp^+ LB agar plate, a portion of the colony was resuspended in 50 µl of distilled water and heated at 95°C for five minutes. Five µl of this solution was added directly as template to a standard PCR reaction containing 1.5 units AmpliTaq® enzyme, 0.1 µM each of T3 and T7 primers, 2.5 mM MgCl2, 10 mM Tris, and 50 mM KCl (pH 8.3). The reactions were amplified in an Ericomp thermocycler with cycling parameters of (95°C, 5 minutes) x 1, (95°C, 10 seconds; 60°C, 30 seconds, 72°C, 2 minutes) x 35, 72°C, 5 minutes.

VII: DNA SEQUENCING

In cycle sequencing, a thermostable DNA polymerase is used for multiple rounds of DNA synthesis at high temperature. A set of four reactions generates sequence data from one strand (forward or reverse) of the template DNA. Each reaction contains a base-specific,

chain terminating, dideoxynucleotide (i.e. ddATP, ddCTP, ddGTP, ddTTP), deoxynucleotides and [α-^{32}P]dATP (10mCi/ml). The primers used for sequencing were M13, forward and reverse primers. Like T3 and T7 oprimers, M13 primers flank the multiple cloning site and encompass the insert. The polymerase extends the chain until a dideoxy nucleotide is incorporated which results in chain termination. The resulting radiolabeled nucleotides are then loaded in adjacent lanes of a polyacrylamide sequencing gel. Both forward and reverse strands were sequenced for Figure 2, part VI.

Sequencing reaction

All reagents were used as supplied by the manufacturer, Epicentre Technologies. A premix consisting of 2.5 µl buffer, 1.5 µl forward or reverse primer (1.5mM), 1 µl [α-^{32}P]dATP (10 mCi/ml), 200 ng DNA was made upto a volume of 17 µl with distilled H_2O. The solution was mixed and 4 µl was aliquoted into four tubes each containing 2 µl of G, A, T, or C termination mix. Each reaction was overlayed with mineral oil and the reactions were cycled in a Ericomp thermocycler with parameters of 95°C, 5 minutes x 1, (95°C, 30 seconds, 70°C, 1 minute) x 30. after cycling 3 µl STOP solution was added to each reaction prior to separation on a sequencing gel.

Sequencing gel

A sequencing apparatus from Jordan Scientific Company (Bloomington, IN) was used. A 5% polyacrylamide gel (38:2 acrylamide:bis) containing 8M urea was used. The gel was loaded and run at 2200 V in 1XTBE (0.09 M Tris base, 0.088 M boric acid, 2 mM EDTA) buffer for one and a half hours. The gel was then fixed in 10% acetic acid for two hours with one change of solution, air-dried and exposed to X-OMAT film.

VIII: SOUTHERN HYBRIDIZATION AND INSERT LABELING

In Southern hybridization, DNA fragments are denatured immobilized on a solid support medium such as nitrocellulose or nylon. The process of immobilization is referred to as "vacuum blotting". The probe is labeled, denatured to make it single stranded and hybridized to the membrane for several hours at a particular stringency. Stringency is a function of both, the ionic strength of the medium, and the temperature of hybridization. The blot is then washed in order to remove background hybridization, dried, and exposed to X-ray film. Southern hybridization was carried out according to Landau-Ellis et al., [1991] and inserts were labeled using the Mega prime kit from Amersham International,

Buckinghamshire, England.

Labeling insert in low melting point agarose

The plasmid clone containing the insert (DAF233 and DAF220) was first purified using the min-prep procedure for isolating plasmid DNA according to Sambrook et al., [1989] with no modifications. The insert was amplified as described in the preceeding section. The amplified insert was resolved on a 1.8% low melting point agarose gel. The gel was stained briefly with ethidium bromide and the insert band was excised from the gel with the minimum of excess agarose and transferred to a pre-weighed 1.5 ml microcentrifuge tube. The tube was weighed again with the gel slice and the weight of the slice determined. The slice was either used immediately or stored at -20°C until further use. The frozen slice was thawed and water was added to ratio of 3 ml per gram of gel and placed in a boiling water bath for 5 minutes to melt the gel and denature the DNA.

All reagents for labeling were used as supplied by the manufacturer. An appropriate volume containing 25 ng of DNA was used for labeling. To this volume, 5 µl of primer at a concentration supplied by the manufacturer was added and the volume was made up to 50 µl. The DNA was denatured by heating to 95-100°C for 5 minutes in a boiling water bath. The tube was spun briefly in a microcentrifuge. While keeping the tube at room temperature, 4 µl of 300 µM nucleotides (omitting dATP), 5 µl reaction buffer (10 mM Tris HCl pH 7.5, 1 mM 2-mercaptoethanol, 1 mM $MgCl_2$), 5 µl [α-^{32}P]dATP (10 mCi/ml), and 2 µl enzyme (1unit/µl). The mixture was gently mixed, spun briefly, and incubated at 37°C for 10 minutes. The reaction was stopped by the addition of 5 µl of 0.2 M EDTA. To 21 µl of the labeling reaction, was added 20 µl SSC buffer. The sephadex matrix in the columns (Boehringer Mannheim Quick Spin) was mixed and allowed to drain by gravity for approximately 20 minutes. The columns were prespun in a larger centrifuge tube for 2 minutes at 2200 rpm using a 1.5 ml microcentrifuge tube to collect excess buffer. The diluted reaction mix (41 µl) was added to the top of the column (with a fresh microcentrifuge tube) taking care not to add to the sides of the column and spun for 4 minutes at 2400 rpm. One µl of the eluant was measured in a scintillation counter to quantify ^{32}P incorporation. The labeled probe was heated to 95°C for 5 minutes before adding to the hybridization solution.

Vacuum blotting

The process of vacuum blotting used in this dissertation is an alkaline transfer method. The

apparatus used for the transfer is a Pharmacia/LKB Vacugene vacuum blotter. The membrane used for blotting was Zeta-Probe (BioRad), a positively charged membrane that binds to negatively charged DNA. To begin the procedure, a plastic window was cut which was approximately 2 cm smaller (length x width) than the gel. A piece of nylon membrane slightly smaller than the window was cut. The membrane was handled gently to avoid any kinks and pre-wet with deionized water. The membrane was placed inside the window and smoothed gently to remove air bubbles. The gel was slid on to the membrane with the wells above the membrane. Care was taken to slide the gel correctly the first time as once the gel is on the memebrane, it becomes difficult to change its position. The wells of the gel should not be blotted. Hydrochloric acid (0.5 M) was added to the surface of the gel while adjusting the vacuum to 20 cm H_2O pressure. The vacuum was turned up to 30 cm H_2O for 10 minutes or until the blue dye (loading dye) turns yellow. This step is necessary for depurinating the DNA fragments. The vacuum was reduced to 20 cm H_2O. The HCl was removed and discarded by pipetting. To the surface of the gel was added, 0.4 M NaOH and the vacuum pressure was increased to 40 cm H_2O. Sodium hydroxide was continued to be added to a depth of 1.5 times the thickness of the gel. Transfer was carried out for 1 hour. The function of sodium hydroxide is to denature the DNA and allow it to bimd to the membrane as single stranded DNA. After the NaOH was removed the membrane was rinsed with 2X SSC until the pH was 7.0. The membrane was placed on blotting paper until the water droplets evaporated. The membrane was then placed in a UV crosslinker from Stratagene for approximately 1 minute (pre calibrated auto crosslink 1200 $\mu J/cm^2$. This step is necessary to crosslink the DNA more tightly to the membrane and is especially important if the membrane will be probed more than once.

Hybridization

The membrane was placed in a heat seal bag and prehybridization solution (1 mM EDTA, 0.5 M $NaHPO_4$ pH 7.2, 7% SDS) was added to it. The bubbles were smoothed out and the bag sealed. The amount of solution depends on the size of the membrane, optimum is 150 μl per cm^2 of membrane area. The membrane was prehybridized for 30 minutes. Prehybridization is necessary to block the positively charged membrane and allow the probe to bind only to the DNA. The prehybridization was replaced with hybridization solution (1 mM EDTA, 0.5 M $NaHPO_4$ pH 7.2, 7% SDS). The labeled probe is heated to 95°C for 10 minutes, spun for 5 minutes in a microcentrifuge and kept on ice. The probe is then added to the hybridization solution. The hybridization solution is placed in a pan of H_2O in a shaking water bath and allowed to shake for at least 4 hours (can be overnight).

After hybridization, the membranes were washed to remove any non-specific binding of probe. Only those fragments with significant homology to the probe remained radiolabelled. Hybridizing and washing at higher temperatures is more stringent. For washing, the membranes were placed in a pan with wash I (1mm EDTA, 40 mm NaHPO4 pH 7.2, 5% SDS) for approximately 5 minutes. The solution was replaced with fresh hybridization solution, and the pan was shaken gently in a 60°C water bath for 60 minutes with one change of solution. The membrane was transfered to a fresh pan with wash II (1 mM EDTA, 40 mM $NaHPO_4$ pH 7.2, 1% SDS) with gentle shaking for 1 hour with one change of solution. The membrane was allowed to air dry (not completely) on 3MM blotting paper. The membrane was sealed in a heat sealed bag. The membranes were placed in a casette, exposed to X-ray film and placed at -70°C.

Development of autoradiogram

The casette was removed from the freezer and allowed to warm to room temperature. The film was removed from the casette and placed in developer solution for 4 minutes. The film was rinsed quickly in water and put in fixative for 2-3 minutes. It was then rinsed for 20-30 minutes under running water and then dried.

BIBLIOGRAPHY

Ahn S, Tanksley SD: Comparative linkage maps of the rice and maize genomes. Proc Natl Acad Sci USA 90: 7980-7984 (1993).

Akkaya MS, Bhagwat AA, Cregan PB: Length polymorphism of simple sequence repeat DNA in soybean. Genetics 132: 1131-1139 (1992).

Akopyanz N, Bukanov NO, Westblom TU, Kresovich S, Berg DE: DNA diversity among clinical isolates of *Helicobacter pylori* detected by PCR-based RAPD fingerprinting. Nucl Acids Res 20: 5137-5142 (1992).

Allen FL: Usefulness of plant genome mapping to plant breeding. In Gresshoff, PM (ed), Plant Genome Analysis. Current Topics in Plant Molecular Biology. pp. 11-18. CRC Press Inc., Boca Raton, Florida, USA (1994).

Allen RC, Graves G, Budowle B: Polymerase chain reaction amplification products separated on rehydratable polyacrylamide gels stained with silver. Biotechniques 7: 736-744 (1989).

Appelbaum E: The *Rhizobium/Bradyrhizobium*-Legume symbiosis. In Gresshoff, PM (ed), Molecular Biology of Symbiotic Nitrogen Fixation. pp. 132-134. CRC Press Inc, Boca Raton, Florida (1990).

Arumanagathan K, Earle ED: Nuclear DNA content of some important plant species. Pl Mol Biol Rep 9: 229-241 (1991).

Ashikawa I, et al: Physical mapping and chromosome mapping. Rice Genome 1: 8-9 (1992).

Ayliffe MA, Lawrence GL, Ellis JG, Pryor AJ: Heteroduplex molecules formed between allelic sequences cause nonparental bands. Nucl Acids Res 22: 1632-1636 (1994).

Barker DG, Bianchi S, Blondon F, Dattée Y, Duc G, Essad S, Flament P, Gallusci G, Génier G, Guy P, Muel X, Tourneur J, Dénarié J, Huget T: *Medicago truncatula*, a model plant for studying the molecular genetics of the *Rhizobium*-legume symbiosis. Plant Mol Biol Rep 8: 40-49 (1990).

Bassam BJ, Bentley S: DNA fingerprinting using Arbitrary Primer Technology (APT): A tool or a torment. Aust Biotech 4: 232-236 (1994).

Bassam BJ, Caetano-Anollés G, Gresshoff PM: Fast and sensitive silver staining of DNA in polyacrylamide gels. Anal Biochem 196: 80-83 (1991).

Bassam BJ, Caetano-Anollés G, Gresshoff PM: DNA amplification fingerprinting of bacteria. Appl Microbiol Biotechnol 38: 70-76 (1992).

Bassam BJ, Caetano-Anollés G, Gresshoff PM: Silver staining of DNA in polyacrylamide gels. Appl Biochem Biotechnol 112: 181-188 (1993).

Bauer P, Crespi MD, Szecsi J, Allison LA, Schultze M, Ratet P, Kondorosi E, Kondorosi A: Alfalfa Enod12 genes are differentially regulated during nodule development by Nod factors and *Rhizobium* invasion. Pl Physiol 105: 585-592 (1994).

Bauer WD: Infection of legumes by *Rhizobia*. Ann Rev Pl Physiol 32: 407-432 (1981).

Beidler JL, Hilliard PR, Rill RL: Ultrasensitive staining of nucleic acids with silver. Anal Biochem 126: 374-380 (1982).

Bender G, Rolfe BG: A rapid plant assay for the *Parasponia-Rhizobium* symbiosis. Plant Science 38: 135-140 (1985).

Bent AF, Kunkel BN, Dahlbeck D, Brown KL, Schmidt R, Giraudat J, Leung J, Staskawacz BJ: *RPS2* of *Arabidopsis thaliana*: A leucine rich repeat class of plant disease resistance genes. Science 265: 1856-1860 (1994).

Bhuvaneswari TV: Recognition mechanisms and infection process in legumes and *Rhizobium*. Econ-Bot 35: 204-223 (1981).

BioVentures: BioVentures manual, Murfreesboro, TN, USA.

Bisseling T, Franssen H, Govers F, Horvath B, Moerman M, Scheres B, de Weil CV, Yang W-C: Early nodulins in pea and soybean development. In Hennecke, H, Verma, DPS (eds), Advances in Molecular Genetics of Plant-Microbe Interaction, vol I. Current Plant Science and Biotechnology in Agriculture. pp. 300-303. Kluwer Academic Publishers, Netherlands (1991).

Blackhall NW, Hammatt N, Davey MR: Analysis of variation in the DNA content of Glycine species. Soybean Genetics Newsletter 18: 194-200 (1991).

Blauenfeldt J, Joshi PA, Gresshoff PM, Caetano-Anollés G: Nodulation of white clover (*Trifolium repens*) in the absence of *Rhizobium*. Protoplasma 179: 106-110 (1994).

Bohlool B: Introduction to nitrogen fixation in agriculture and industry: Contribution of BNF to sustainability of agriculture. In Gresshoff, PM, Roth, LE, Stacey, G, Newton, WE (eds), Nitrogen Fixation: Acheivements and Objectives. Proccedings of the 8th International Congress on Nitrogen Fixation, Knoxville, TN, USA. pp. 613-615. Chapman Hall, New York, London (1990).

Bonierbale MW, Plaisted RL, Tanksley SD: RFLP maps based on a common set of clones reveal modes of chromosomal evolution in potato and tomato. Genetics 120: 1095-1103 (1988).

Botstein D, White RL, Skolnick M, Davis RW: Construction of a genetic map in man using restriction fragment length polymorphisms. Am J Hum Genet 32: 314-331 (1980).

Boulter D: The use of amino acid sequence data in phylogenetic studies with special reference to plant proteins. In Bisby, FA, Vaughan, JG, Wright, CA (eds), Chemosystematics, Principles and Practice. pp. 235-240. Academic Press, London (1980).

Boutin S, Olson T, Yu Z-H, Young N, Shoemaker R, Vallejos E: Genome conservation among legume genera detected with DNA markers. Abstract in the proceedings of the Molecular & Cellular Biology of the Soybean, 5th Biennial Conference, University of Georgia. p. 13. (1994).

Burke DT, Carle GF, Olson M: Cloning of large segments of exogenous DNA into yeast by means of artificial chromosome vectors. Science 236: 806-812 (1987).

Burr B, Burr FA: Recombinant inbreds for molecular mapping in maize: theoretical and practical considerations. Trends Genet 7: 55-60 (1991).

Cabané M, Vincens P, Boudet AM: Protein synthesis at low temperatures in two soybean cultivars differing by their cold sensitivity. Physiol Plantarum 85: 573-580 (1992).

Caetano-Anollés G, Bauer WD: Feedback regulation of nodule formation in alfalfa. Planta 175: 546-547 (1988a).

Caetano-Anollés G, Bauer WD: Enhanced nodule intitiation on alfalfa by wild type *Rhizobium meliloti* co-inoculated with *nod* gene mutants and other bacteria. Planta 175: 546-557 (1988b).

Caetano-Anollés G, Gresshoff PM: Early induction of feedback regulatory responses governing nodulation in soybean. Plant Science 71: 169-181 (1990).

Caetano-Anollés G, Gresshoff PM: Alfalfa controls nodulation during the onset of *Rhizobium*-induced cortical cell division. Plant Physiol 95: 366-373 (1991a).

Caetano-Anollés G, Gresshoff PM: Plant genetic control of nodulation. Ann Rev Microbiol 45: 345-382 (1991b).

Caetano-Anollés G, Bassam BJ, Gresshoff PM: DNA amplification fingerprinting: A strategy for genome analysis. Pl Mol Biol Rep 9: 294-307 (1991a).

Caetano-Anollés G, Bassam BJ, Gresshoff PM: DNA amplification fingerprinting using very short arbitrary oligonucleotide primers. Bio/technology 9: 553-556 (1991b).

Caetano-Anollés G, Joshi PA, Gresshoff PM: Spontaneous nodules induce feedback suppression of nodulation in alfalfa. Planta 183: 77-82 (1991c).

Caetano-Anollés G, Bassam BJ, Gresshoff PM: Primer-template interactions during DNA amplification fingerprinting with single arbitrary oligonucleotides. Mol Gen Genet 235: 157-165 (1992a).

Caetano-Anollés G, Bassam BJ, Gresshoff PM: Fingerprinting genomes with arbitrary oligonucleotide primers. Mol Biol Sci Adv 11: 1-6 (1992b).

Caetano-Anollés G, Bassam BJ, Gresshoff PM: Enhanced detection of polymorphic DNA by multiple arbitrary amplicon profiling of endonuclease-digested DNA: identification of markers tightly linked to the supernodulation locus in soybean. Mol Gen Genet 241: 57-64 (1993).

Caetano-Anollés G: MAAP: a versatile and universal tool for genome analysis. Pl Mol Biol 25: 1011-1026 (1994).

Caetano-Anollés G, Gresshoff PM: DNA amplification fingerprinting using arbitrary mini-hairpin oligonucleotide primers. Bio/technology 12: 619-623 (1994).

Caldwell BE: Inheritance of a strain-specific ineffective nodulation in soybean. Crop Sci 6: 427-428 (1966).

Callahan LM, Weaver KR, Caetano-Anollés G, Gresshoff, PM: DNA Fingerprinting of turfgrasses. Int Turfgrass Soc Res J 761-767 (1993).

Calvert HE, Pence MK, Pierce M, Malik NS, Bauer WD: Anatomical analysis of the development and distribution of *Rhizobium* infections in soybean roots. Can J Botany 62: 2375-2384 (1984).

Carlson JE, Tulsieram LK, Glaubitz JC, Luk VWK, Kauffeldt C, Rutledge R: Segregation of random amplified DNA markers in F1 progeny of conifers. Theor Appl Genet 83: 194-200 (1991).

Carroll BJ, McNeil DL, Gresshoff PM: Isolation and properties of soybean (*Glycine max*) mutants that nodulate in the presence of high nitrate concentration. Proc Natl Acad Sci USA 82: 4162-4166 (1985a).

Carroll BJ, McNeil DL, Gresshoff PM: A supernodulation and nitrate tolerant symbiotic (nts) soybean mutant. Plant Physiol 78: 34-40 (1985b).

Carroll BJ, McNeil DL, Gresshoff PM: Mutagenesis of soybean (*Glycine max* (L.) Merr.) and the isolation of non-nodulating mutants. Plant Sci 109-114 (1986).

Carroll BJ, Mathews A: Nitrate inhibition of nodulation in legumes. In Gresshoff, PM (ed), The Molecular Biology of Symbiotic Nitrogen Fixation. pp. 159-180. CRC Press, Boca Raton (1990).

Chuck G, Robbins T, Nijjar C, Ralston E, Courtney-Gutterson N, Dooner HK: Tagging and cloning of a petunia flower color gene with the maize transposable element *Activator*. Plant Cell 5: 371-378 (1993).

Chumakov I: Continuum of overlapping clones spanning the entire human chromosome 21q. Nature 359: 380-387 (1992).

Clegg MT: Chloroplast gene sequences and the study of plant evolution. Proc Natl Acad Sci USA 90: 363-367 (1993).

Cohen AS, Najarian DR, Karger BL: Separation and analysis of DNA sequence reaction products by capillary gel electrophoresis. J Chromatography 516: 49-60 (1990).

Cooper DN, Schmidtke J: Molecular genetic approaches to the analysis and diagnosis of human inherited disease: an overview. Ann Med 24: 29-42 (1992).

Costa GL, Weiner MP: Polishing with T4 or *Pfu* polymerase increases the efficiency of cloning of PCR fragments. Nucl Acids Res 22: 242 (1994).

Cram WJ. In Rennenberg, H, Brunold, CH, Dekok, JL, Stulen, I (eds), Uptake and transport of sulfate in sulfur nutrition in higher plants. Fundamental Environmental and Agricultural Aspects. SPB Academic Publishing, Netherlands (1990).

Crawford DJ: Electrophoretic studies of Rubisco contributions to plant systematics and phylogeny. Plant Molecular Systematics Macromolecular Approaches. pp. 51-64. John Wiley & Sons Inc, New York, (1990).

Cregan PB, Akayya MS, Bhagwat AA, Lavi U, Rrongwen J: Length polymorphisms of simple sequence repeat (SSR) DNA as molecular markers in plants. In Gresshoff, PM (ed), Plant Genome Analysis. Current Topics in Plant Molecular Biology. pp. 47-56. CRC Press Inc., Boca Raton, Florida, USA (1994).

Crespi MD, Jurkevitch E, Poiret M, d'Aubenton-Carafa Y, Petrovics G, Kondorosi E, Kondorosi A: *ENOD40*, a gene expressed during nodule organogenesis, codes for a non-translatable RNA involved in plant growth. EMBO J 13: 5099-5112 (1994).

Dazzo FB, Brill WJ: Regulation by fixed N_2 of host-symbiont recognition in the *Rhizobium*-clover symbiosis. Plant Physiol 62: 1814-1817 (1978).

Dellaporta SL, Wood J, Hicks JB: A plant DNA mini-preparation. Version II. Plant Mol Biol Rep 1: 19-21 (1983).

Delourme R, Boucherean A, Hubert N, Renard M, Landry BS: Identification of RAPD markers linked to a fertility restorer gene for the *Ogura* radish cytoplasmic male sterility gene of rapeseed (*Brassica napus* L.). Theor Appl Genet 88: 741-748 (1994).

Delves AC, Mathews A, Day DA, Carter AS, Gresshoff PM: Regulation of the soybean-*Rhizobium* nodule symbiosis by shoot and root factors. Pl Physiol 82: 588-590 (1986).

Delves AC, Higgins A, Gresshoff PM: Supernodulation in interspecific grafts between *Glycine max* (soybean) and *Glycine soja*. J Pl Physiol 128: 473-478 (1987a).

Delves AC, Higgins A, Gresshoff PM: Shoot control of supernodulation in a number of mutant soybeans (*Glycine max* (L.) Merr.). Aust J Plant Physiol 14: 689-694 (1987b).

Delves AC, Carroll BJ, Gresshoff PM: Genetic analysis and complementation studies on a number of mutant supernodulating soybeans. J Genet 67: 1-8 (1988).

Devos KM, Gale MD: The use of random amplified polymorphic DNA markers in wheat. Theor Appl Genet 84: 567-572 (1992).

Devos KM, Atkinson MD, Chinoy CN, Liu CJ, Gale MD: Chromosomal rearrngement in rye genome relative to that of wheat. Theor Appl Genet 83: 931-939 (1992).

Diers BW, Mansur L I, J Shoemaker R: Mapping of the *Phytophthora* root rot resistance loci in soybean with restriction fragment length polymorphism markers. Crop Sci 32: 377-383 (1992a).

Diers BW, Keim P, Fehr WR, Shoemaker RC: RFLP analysis of soybean seed protein and oil content. Theor Appl Genet 83: 608-612 (1992b).

Dietrich S, Katz H, Lincoln SE, Shin HS, Friedman J, Dracopoli NC, Lander E-S: A genetic map of the mouse suitable for typing inter-specific crosses. Genetics 131: 423-447 (1992).

Dobereiner J, Day JM: Associative symbiosis in tropical grasses: Characterization of microorganisms and dinitrogen fixing sites. In Newton, WE, Nyman, CJ (eds), Proc. 1st symposium. Washington State University Press, Pullman, USA, (1976).

Donis-Keller H, Green P, Helms C, Cartinhour S, Weiffenbach B, Stephens K, Keith TP, Bowden DW, Smith DR, Lander ES et al: A genetic map of the human genome. Cell 51: 31 (1987).

Durkin AS, Maglott DR, Nierman WC: Chromosomal assignment of 38 human brain-expressed sequence tags (ESTs) by analyzing fluorescently-labeled PCR products from hybrid cell panels. Genomics 14: 808-810 (1992).

Dweikat I, Mackenzie S, Levy M, Ohm H: Pedigree assessment using RAPD-DGGE in cereal crop species. Theor Appl Genet 85: 497-505 (1993).

Eady RR, Smith BE: Physico-chemical properties of nitrogenase and its components. In Hardy, RWF, BoMomley, F, Burns, RC (eds), A treatise on dinitrogen fixation, second edition. pp. 399-490. Wiley Interscience, New York (1979).

Edwards MD, Stuber CW, Wendel JF: Molecular-marker-facilitated investigations of Quantitative-Trait Loci in maize. I. numbers, genomic distribution and types of gene action. Genetics 116: 113-125 (1987).

Ellsworth DL, Rittenhouse KD, Honeyartt LR: Artifactual variation in randomly amplified polymorphic DNA banding patterns. Bio/techniques 14: 214-218 (1992).

Eskew DL, Caetano-Anollés G, Bassam BJ, Gresshoff PM: DNA amplification fingerprinting of the *Azolla-Anabena* symbiosis. Pl Mol Biol 21: 363-373 (1993).

Fearn JC, LaRue TA: An altered constitutive peptide in *sym5* mutants of *Pisum sativum* L. Plant Mol Biol 14: 207-216 (1990).

Figueiras AM, Zaragoza C, Gallego FJ, Benito C: NADH dehydrogenase: a new molecular marker for homeology group 4 in *Triticeae*. a map of the 4RS chromosome in rye. Theor Appl Genet 83: 169-172 (1991).

Flensgrud R, Kobro G: A method for two-dimensional electrophoresis of proteins from green plant tissues. Anal Biochem 177: 33-36 (1989).

Foote S, Vollrath D, Hilton A, Page DC: The human Y chromosome: overlapping DNA clones spanning the euchromatic region. Science 258: 60-66 (1992).

Franssen H, van-Lierop MJ, van-Lammeren A, van-Lammeren A, Bisseling T: The early nodulin transcript ENOD2 is located in the nodule parenchyma (inner cortex) of pea and soybean root nodules. EMBO J 9: 1-7 (1990).

Fritsch P, Hanson MA, Spore CD, Pack PE, Rieseberg LH: Constancy of RAPD primer amplification strength among distantly related taxa of flowering plants. Pl Mol Biol Rep 11: 10-20 (1993).

Fukuoka S, Hosaka K, Kamijuna O: Use of random amplified polymorphic DNAs (RAPDs) for identification of rice accessions. Jpn J Genet 67: 243-252 (1992).

Funke R, Kolchinsky A, Gresshoff PM: Physical mapping of a region in the soybean (*Glycine max*) genome containing duplicated sequences. Pl Mol Biol 22: 437-446 (1993).

Funke RP, Kolchinsky A, Gresshoff, PM: High EDTA concentrations cause entrapment of small DNA molecules in the compression zone of pulsed field gels, resulting in smaller than expected insert sizes in YACs prepared from size selected DNA. Nucl Acids Res 22: 2708-2709 (1994).

Ganal MW, Lapitan NLY, Tanksley SD: A moelcular and cytogenetic survey of major repeated DNA sequences in tomato (*Lycopersicon esculentum*). Mol Gen Genet 213: 262-268 (1988).

Ganal MW, Lapitan NLY, Tanksley SD: Macrostructure of the tomato telomeres. Plant Cell 3: 87-94 (1991).

Gebhardt C, Ritter E, Barone A, Debener T, Walkemeier B, Schachtschabel U, Kaufman H, Thompson RD, Bonierbale MW, Ganal MW, Tanksley SD, Salami F: RFLP maps of potato and their alignment with the homeologous tomato genome. Theor Appl Genet 83: 49-57 (1991).

Gerahty N, Caetano-Anollés G, Joshi PA, Gresshoff PM: Distribution of *Bradyrhizobium*-induced cell division in soybean. In Gresshoff, PM, Roth, LE, Stacey, G, Newton, WE (eds), Nitrogen Fixation: Acheivements and Objectives. Proccedings of the 8th International Congress on Nitrogen Fixation, Knoxville, TN, USA. p. 737. Chapman Hall, New York, London (1990).

Gibson AH, Harper JE: Nitrate effect on nodulation of soybean by *Rhizobium japonicum*. Crop Sci 25: 497-501 (1985).

Giovanelli J, Mudd S, Datko AH: Sulfur amino acids in plants. In Miflin, BJ (ed), The biochemistry of plants. vol 5 A comprehensive treatise. Amino acids and derivatives. pp. 454-506. Academic Press, London, UK (1980).

Goldberg RB: DNA sequence organization in the soybean plant. Biochem Gen 16: 1-12 (1978).

Goldberg RB: Regulation of plant gene expression. Phil Trans R Soc., London, Ser. B. 314: 434-453 (1988).

Goodman HM, Giraudat J, Hauge BM, Valon C, Smalle J, Parcy F: Isolation of the *Arabidopsis AB13* gene by postional cloning. Plant Cell 4: 1251-1261 (1992).

Gremaud MF, Harper JE: Selection and initial characterization of partially nitrate-tolerant nodulation mutants of soybean. Plant Physiol 89: 169-173 (1989).

Gresshoff PM, Delves AC: Plant genetic approaches to symbiotic nodulation and nitrogen fixation in egumes. In Blonstein, AD, King, PJ (eds), Plant Gene Research III. A Genetical Approach to Plant Biochemistry. pp. 159-206. Springer Verlag, Wein (1986).

Gresshoff PM, Mathews A, Krotzky AJ, Olsson JE, Carroll BJ, Delves AC, Kosslak R, Appelbaum ER, Day DA: Molecular genetics of plant-microbe interactions. Amer Phytopath Soc St. Paul 364-369 (1988a).

Gresshoff PM, Krotzky A, Mathews A, Day DA, Schuller KA, Olsson J, Delves AC, Carroll BJ: Suppression of the symbiotic supernodulation symptoms of soybean. J Plant Physiol 132: 417-423 (1988b).

Gresshoff PM, Olsson JE, ZhiZheng L, Caetano-Anollés G: Root-shoot interaction in the control of nodulation in soybean. Current topics in Plant Biochemistry and Physiology 8: 125-139 (1989).

Gresshoff PM: Plant function in nodulation and nitrogen fixation in legumes. In Palacios, R, Mora, J, Newton, WE (eds), New Horizons in Nitrogen Fixation. Current Plant Science and Biotechnology in Agriculture. pp. 31-42. Kluwer Publishing, Netherlands (1993).

Gresshoff PM and Landau-Ellis D: Molecular mapping of soybean nodulation genes. In Gresshoff, PM (ed), Plant Genome Analysis. Current Topics in Plant Molecular Biology. pp. 97-112. CRC Press Inc., Boca Raton, Florida, USA (1994).

Gresshoff PM, MacKenzie AK: Low experimental variability of DNA profiles generated by arbitrary primer based amplification (DAF) of soybean. Chinese J Bot 6: 1-6 (1994).

Grierson D, Covey S: The plastome and chloroplast biogenesis. In Grierson, D, Covey, S (eds), Plant Molecular Biology, 2nd edn. pp. 47-77. Blackie and Son Ltd, Chapman & Hall, New York, NY, USA (1988).

Guinel FC, LaRue TA: Light microscopy study of nodule initiation in *Pisum sativum* L. cv. Sparkle and its low nodulation mutant E2 (*sym* 5). Plant Physiol 97: 1206-1211 (1991).

Gurley WB, Hepburn AU, Key JL: Sequence organization of the soybean genome. Biochim Biophys Acta 561: 167-183 (1979).

Guszczynski T, Chrambach A: Electrophoretic separation of *S. pombe* chromosomes in polyacrylamide solutions using a constant field. Biochem Biophys Res Commun 179: 482-486 (1991).

Haley SD, Miklas PN, Stavely JR, Byrum J, Kelly JD: Identification of RAPD markers linked to a major rust resistance gene block in common bean. Theor Appl Genet 86: 505-512 (1993).

Haley SD, Afandor LK, Miklas PN, Stavely JR, Kelly JD: Heterogenous inbred populations are useful as sources of near-isogenic lines for RAPD marker utilization. Theor Appl Genet 85: 337-342 (1994).

Halward S, Stalker T, LaRue E, Kochert G: Use of single primer DNA amplification in genetic atudies of peanut (*Arachis hypogaea* L.). Pl Mol Biol 18: 315-325 (1992).

Hamsey KR, Zimmer EA: Ribosomal RNA as a phylogenetic tool in plant systematics. In Soltis, PS, Soltis, DE, Doyle, JJ (eds), Molecular systematics in plants. pp. 50-91. Chapman and Hall, New York, NY, USA (1992).

Handberg K, Stougaard J: *Lotus japonicus*, an autogamous, diploid legume species for classical and molecular genetics. Plant J 2: 487-496 (1992).

Hansen J, Jorgensen J-E, Stougaard J, Marcker KA: Hairy roots-a short cut to transgenic root nodules. Plant Cell Rep 8: 12-15 (1989).

Hellriegel H, Beilageheft Z: Ver Rübenzucker-ind D Reiches (1886).

Henderson ER, Blackburn EH: An overhanging 3' terminus is a conserved feature of telomeres. Mol Cell Biol 9: 345-348 (1989).

He S, Ohm H, Mackenzie S: Detection of DNA sequence polymorphisms among wheat varieties. Theor Appl Genet 84: 573-578 (1992).

Heun M, Kennedy AE, Anderson JA, Lapitan NLV, Sorrels ME, Tanksley SD: Construction of a restriction fragment length polymorphism map for barley (*Hordeum vulgare*). Genome 34: 437-447 (1991).

Hinson K: Nodulation response from nitrogen applied to soybean half root systems. Agronomy J 67: 799-804 (1975).

Hirsch AM, Bhuvaneswari TV, Torrey JG, Bisseling T: Early nodulin genes are induced in alfalfa root outgrowths elicited by auxin transport inhibitors. Proc Natl Acad Sci USA 86: 1244-1248 (1989).

Hirsch AM: Developmental biology of legume nodulation. New Phytol 122: 211-237 (1992).

Hnetkovsky NK, Doubler TW, Gibson PT, Lightfoot D: RFLP and RAPD markers linked to genes affecting soybean sudden death syndrome response. Abstract in the proceedings of the Molecular & Cellular Biology of the Soybean, 5th Biennial Conference, University of Georgia. p. 9. (1994).

Hochstrasser DF, Harrington MG, Hochstrasser AC, Miler MJ, Merril CR: Methods for increasing the resolution of two-dimensional protein electrophoresis. Anal Biochem 173: 424-435 (1988).

Hoeffer Scientific Instruments: Instr. manual for TKO100 Dedicated Minifluometer. San Fransisco, USA.

Hsiao K: Exonuclease III induced ligase-free directional subcloning of PCR products. Nucl Acids Res 21: 5528-5529 (1993).

Hubby JL, Lewontin RC: A molecular approach to the study of genetic heterozygosity in natural populations. I: The number of alleles at different loci in *Drosophila pseudoobscura*. Genetics 54: 577-594 (1966).

Huffaker RC: Biochemistry and Physiology of Leaf Proteins. In Boulter, D, Parthier, B (eds), Nucleic acids and protein in plants I: Structure, Biochemistry and Physiology of proteins. Encyclopedia of Plant Proteins, vol 14A. p. 371. (1982).

Hunt JG, Page RE Jr: Patterns of inheritance with RAPD molecular markers reveal novel types of polymorphism in the honey bee. Theor Appl Genet 85: 15-20 (1992).

Ichihara Y, Korosawa Y: Construction of new T vectors for direct cloning of PCR products. Gene 130: 153-154 (1993).

Inoue T, Zhong HS, Miyao A, Ashikawa I, Monna L, Fukuoka S, Miyadera N, Nagamura Y, Kurata N, Sasaki T, Minobe Y: Sequence-tagged-sites (STSs) as standard landmarks in the rice genome. Theor Appl Genet 89: 728-734 (1994).

Jack T, Brockman LL, Meyerowitz EM: The homeotic gene *APETALA3* of *Arabidopsis thaliana* encodes a MADS box and is expressed in petals and stamens. Cell 68: 683-697 (1992).

Jacobsen E: Modification of symbiotic interaction of pea (*Pisium sativum* L.) and *Rhizobium leguminosarum* by induced mutations. Plant Soil 82: 427 (1984).

James Jr. DW, Lim E, Keller J, Plooy I, Ralston E, Dooner H: Directed tagging of the *Arabidopsis FATTY ACID ELONGATION 1* (*FAE1*) gene with the maize transposon activator. Plant Cell 7: 309-315 (1995).

Jeffreys AJ, Wilson V, Thern SL: Hypervariable "minisatellite" regions in human DNA. Nature 314: 67-73 (1985).

Jeffreys AJ, Wilson V, Thern SL: Individual-specific fingerprints of human DNA. Nature 314: 76-79 (1985).

Johns MA: Comparison of RAPD maps for two sets of maize recombinant inbred lines. Abstract in Plant Genome I Conference, Official Program and Abstracts. p. 32. San Diego, CA, (1992).

Joshi PA, Caetano-Anollés G, Graham ET, Gresshoff PM: Ultrastructure of transfer cells in spontaneous nodules of alfalfa. Protoplasma 172: 64-76 (1993).

Junguo H, Quiros CF: Identification of brocolli and cauliflower cultivars with RAPD markers. Pl Cell Rep 10: 505-511 (1991).

Kasha RJ, Sadasivai RS: Genome relationships between *Hordeum vulgare* L. and *Hordeum bulbosum*. Chromosoma 35: 264-287 (1971).

Kazan K, Manners JM, Cameron DF: Genetic variation in agronomically important species of *Stylosanthes* determined using random amplified polymorphic DNA markers. Theor Appl Genet 85: 882-888 (1993).

Keim P, Shoemaker RC: Construction of a random recombinant DNA library that is primarily single copy sequence. Soybean Genetics Newsletter 15: 147-148 (1988).

Keim P, Shoemaker RC, Palmer RG: Restriction fragment length polymorphism in soybean. Theor Appl Genet 77: 786-792 (1989).

Keim P, Diers BW, Shoemaker RC: Genetic analysis of soybean hard seededness with molecular markers. Theor Appl Genet 79: 465-469 (1990).

Keim P, Diers BW, Olson TC, Shoemaker RC: RFLP mapping in soybean: association between marker loci and variation in quantitative traits. Genetics 126: 735-742 (1990).

Kere J, Nagaraja R, Mumm S, Ciccodicola A, D'Urso M, Schlessinger D: Mapping human chromosomes by walking with sequence-tagged sites from end fragments of yeast artificial chromosome inserts. Genomics 14: 241-248 (1992).

Kesseli R, Witsenboer H, Stanghellini M, Vandermark G, Michelmore R: Recessive resistance to *Plasmopara lactucae-radicis* maps by bulked segregant analysis to a cluster of dominant disease resistance genes in lettuce. Mol Pl Microbe Inter 6: 722-728 (1993).

Kirby LT. DNA Fingerprinting: An Introduction. pp. 1-5. Stockton Press, New York, NY (1990).

Klein-Lankhorst RM, Vermunt A, Weide R, Liharska T, Zabel P: Isolation of molecular markers for tomato (*L. esculentum*) using random amplified polymorphic DNA (RAPD). Theor Appl Genet 83: 108-114 (1991).

Koenig M, Hoffman EP, Bertelson CJ, Monaco AP, Ferner C, Kunkel LM: Complete cloning of the Duchenne Muscular Dystrophy (DMD) cDNA and preliminary genomic organization of the DMD gene in normal and affected individuals. Cell 50: 509-517 (1987).

Kolchinsky A, Gresshoff PM: A major satellite DNA of soybean is a 92-base pairs tandem repeat. Theor Appl Genet in Press: (1995).

Kolchinsky A, Landau-Ellis D, Gresshoff PM: Molecular analysis of a polymorphic DNA region linked to the supernodulating (*nts*) locus of soybean. In review: (1995).

Kolchinsky AM, Funke RP, Gresshoff PM: DAF-Amplified fragments can be used as markers for DNA from pulse field gels. Biotechniques 14: 400-403 (1993).

Kollipara KP, Ahmad P, Burridge JA, Xu SJ, Singh RJ, Hymowitz T: Gene mapping in soybean with primary trisomics. Abstract in the proceedings of the Molecular & Cellular Biology of the Soybean, 5th Biennial Conference, University of Georgia. p. 6. (1994).

Koncz C, Martini N, Mayerhofer R, Koncz-Kalman Z, Korber H, Redei GP, Schell J: High-frequency T-DNA mediated gene tagging in plants. Proc Natl Acad Sci USA 86: 8467-8471 (1989).

Konieczny A and Ausubel FM: A procedure for mapping *Arabidopsis* mutations using codominant ecotype-specific PCR-based markers. Plant J 4: 403-410 (1993).

Kosslak RM, Bohlool BB: Suppression of nodule development of one side of a split root system of soybeans caused by prior inoculation of the other side. Plant Physiol 75: 125-130 (1984).

Kouchi H, Hata S: Isolation and characterization of novel nodulin cDNAs representing at early stages of soybean nodule development. Mol Gen Genet 238: 106-119 (1993).

Kovalic D, Kwak JH, Weisblum B: General method for direct cloning of DNA fragments generated by the polymerase chain reaction. Nucl Acids Res 19: 4560 (1991).

Kreitman M: Nucleotide polymorphisms at the alcohol dehydrogenase locus of *Drosophila melanogaster*. Nature 304: 412-417 (1983).

Kurata N, Moore G, Nagamura Y, Foote T, Yano M, Minobe Y, Gale M: Conservation of genome structure between rice and wheat. Bio/Technology 12: 276-278 (1994).

Ladha JK Pareek RP, So R Becker M: Stem-Nodule symbiosis and its unusual properties. In Gresshoff, PM, Roth, LE, Stacey, G, Newton, WE (eds), Nitrogen Fixation: Acheivements and Objectives. Proccedings of the 8th International Congress on Nitrogen Fixation, Knoxville, TN, USA. p. 633. Chapman Hall, New York London (1990).

Lam BA, Chrambach A: Analytical and preparative gel electrofocussing. In Hames, BD, Rickwood, D (eds), Gel electrophoresis of proteins. p. 181. IRC Press Limited, Oxford, England, (1987).

Landau-Ellis D, Angermüller S, Shoemaker R, Gresshoff PM: The genetic locus controlling supernodulation in soybean (*Glycine max* L.) co-segregates tightly with a cloned molecular marker. Mol Gen Genet 228: 221-226 (1991).

Landau-Ellis D, Gresshoff PM: The RFLP molecular marker closely linked to the supernodulation locus of soybean contains three inserts. Mol Plant Microbe Inter 5: 428-429 (1992).

Lander ES, Green P, Abrahamson J, Barlow A, Daly MJ, Lincoln SE, Newberg L: MAPMAKER: An interactive computer package for constructing primary genetic linkage maps of experimental and natural populations. Genomics 1: 174-181 (1987).

Landers JP: Capillary electrophoresis: pioneering new approaches for biomolecular analysis. Trends Biochem Sci 409-414 (1993).

Lark KG, Weisemann JM, Matthews BF, Palmer R, Chase K, Macalma T: A genetic map of soybean (Glycine max L.) using an intraspecific cross of two cultivars: 'Minsoy' and 'Noir1'. Theor Appl Genet 86: 901-906 (1993).

Lark KG, Orf J, Mansur LM: Epistatic expression of quantitative trait loci (QTL) in soybean [*Glycine max* (L.) Merr.] determined by QTL association with RFLP alleles. Theor Appl Genet 88: 486-489 (1994).

Laurent AM, Marcais M, Roizes G: A rapid and simple method to isolate and characterize highly polymorphic markers from the centromeric region of the human chromosome. Nucl Acids Res 22: 194-199 (1994).

Learn GH, Schaal BA: Population subdivision for ribosomal DNA repaets in *Clematis fremontii*. Evolution 41: 433-438 (1987).

Lee N-G, Stein B, Suzuki H, Verma DPS: Expression of antisense nodulin-35 RNA in *Vigna aconitifolia* transgenic root nodules retards peroxisome development and affects nitrogen availability to the plant. Plant J 3: 599-606 (1993).

Leutwiler LS, Hough-Evans B, Meyerowitz EM. Mol Gen Genet 194: 15 (1984).

Lewontin RC, Hubby JL: A molecular approach to the study of genetic heterozygosity in natural populations. Genetics 54: 595 (1966).

Liang P, Averbonkh L, Pardee AB: Distribution and cloning of eukaryotic mRNAs by means of differential display: refinements and optimization. Nucl Acids Res 21: 3269-3334 (1993).

Lisitsyn NA: Direct isolation of polymorphic markers linked to a trait by genetically directed representational difference analysis. Nature Genet 6: 57-64 (1994).

Liu Y-G, Mitsukawa N, Whittier RF: Rapid sequencing of unpurified PCR products by thermal asymetric PCR cycle sequencing using unlabeled sequencing primers. Nucl Acids Res 21: 3333-3334 (1993).

Liu Z, Schwartz L: An efficient method for blunt-end ligation of PCR products. Biotechniques 12: 28-29 (1992).

Lorens J: Rapid and reliable cloning of PCR products. PCR Meth Appl 1: 140-141 (1991).

Lumpkin TA: Maintenance and cultivation of *Azolla*. In Elkan, GH (ed), Symbiotic Nitrogen Fixation Technology. pp. 55-94. Marcel Dekker, Inc., New York, NY, USA (1987).

MacPherson JM, Eckstein PE, Scoles GJ, Gajadhar AA: Variability of the random amplified polymorphic DNA assay among thermal cyclers and effects of primer and DNA concentrations. Mol Cell Probe 7: 293-299 (1993).

Malik NSA, Calvert HE, Bauer WD: Nitrate induced regulation of nodule formation in soybean. Plant Physiol 84: 266-271 (1987).

Mansur LM, Orf J, Lark KG: Determining the linkage of quantitative trait loci to RFLP markers using extreme phenotypes of recombinant inbreds of soybean (*Glycine max* L. Merr.). Theor Appl Genet 86: 914-918 (1993).

Mao L, Lee DJ, Tockman MS, Erozan VS, Askin F, Sdransky D: Microsatellite alterations as clonal markers for the detection of human cancer. Proc Natl Acad Sci USA 91: 9871-9875 (1994).

Martin GB, Williams JGK, Tanksley S: Rapid identification of markers linked to a *Pseudomonas* resistance gene in tomato by using random primers and near isogenic lines. Proc Natl Acad Sci USA 88: 2336-2340 (1991).

Martin GB, Brommonschenkel SH, Chunwongse J, Frary A, Ganal MW, Spivey R, Wu T, Earle ED, Tanksley SD: Map-based cloning of a protein kinase gene conferring resistance in tomato. Science 262: 1432-1436 (1993).

Martin GB, Frary A, Wu T, Brommonschenkel S, Chunwongse J, Earle ED, Tanksley SD: A member of the tomato *Pto* gene family confers sensitivity to Fenthion resulting in rapid cell death. Pl Cell 6: 1543-1552 (1994).

Martinez E, Romero D, Palacios R: The *Rhizobium* genome. Critical Reviews in Plant Sciences 9: 59-93 (1990).

Mathews A, Carroll BJ, Gresshoff PM: Characterization of non-nodulation mutants of soybean (*Glycine max* (L.) Merr.): *Bradyrhizobium* effect and absence of root hair curling. J Plant Physiol 131: 349-361 (1987).

Mathews A, Carrol BJ, Gresshoff PM: Development of *Bradyrhizobium* infections in a supernodulating and non-nodulating mutant of soybean [*Glycine max* (L.) Merr.]. Protoplasma 150: 40-47 (1989).

Matvienko M, Van de Sande K, Yang WC, Van-Kammen A, Bisseling T, Franssen H: Comparison of soybean and pea ENOD40 cDNA clones representing genes expresses during both early and late stages of nodule development. Pl Mol Biol 26: 487-493 (1994).

Mayrand PE, Corcoran KP, Ziegle JS, Robertson JM, Hoff BL, Kronick MN: The use of fluorescence detection and internal standards to size PCR products automatically. Appl Theor Elec 3: 1-11 (1992).

McCouch SR, Kochert G, Yu ZH, Wang ZY, Khush GS, Coffman WR, Tanksley SD: Molecular mapping of rice chromosomes. Theor Appl Genet 76: 815-829 (1988).

Merril CR, Goldman D, Sedman SA, Ebert MH: Ultrasensitive stain for proteins in polyacrylamide gels shows regional variation in cerebrospinal fluid proteins. Science 211: 1437-1438 (1981).

Merril CR: Silver staining of proteins and DNA. Nature 343: 779-780 (1990).

Messager A: Selection of pea mutants for nodulation and nitrogen fixation. Analysis of the plant genes involved in the legume-*Rhizobium* symbiosis. Organization for Economic Cooperation and Development, Paris. p. 52. (1985).

Meunier J-R, Grimont PAD: Factors affecting reproducibility of random amplified polymorphic DNA fingerprintng. Res Microbiol 144: 373-379 (1993).

Micheli MR, Bova R, Calissano P, D'Ambrosio E: Randomly amplified polymorphic DNA fingerprinting using combinations of oligonucleotide primers. Bio/techniques 15: 388-389 (1993).

Michelmore RW, Paran I, Kesseli RV: Identification of markers linked to disease-resistance genes by bulked segregant analysis: A rapid method to detect markers in specific genomic regions by using segregating populations. Proc Natl Acad Sci USA 88: 9828-9832 (1991).

Miwa T, Sudo K, Nakamura Y, Imai T: Fifty sequence-tagged sites on human chromosome 11. Genomics 17: 211-214 (1993).

Morgante M, Rafalski A, Biddle P, Tingey S, Olivieri AM: Genetic mapping and variability of seven soybean simple sequence repeat loci. Genome 37: 763-769 (1994).

Muehlbauer GJ, Specht JE, Thomas-Compton MA, Staswick PE, Bernard RL: Near-isogenic lines-a potential resource in the integration of conventional and molecular linkage maps. Crop Sci 28: 729-735 (1988).

Muehlbauer GJ, Staswick PE, Specht JE, Graef GL, Shoemaker RC, Keim P: RFLP mapping using near-isogenic lines in the soybean [*Glycine max* (L.) Merr.]. Theor Appl Genet 81: 189-198 (1991).

Mueller PR: Ligation-mediated PCR for genomic sequencing and footprinting. In Ausubel, FM, Brent R, Kingston RE, Moore DD, Seidman JG, Smith JA, Struhl K (eds), Current Protocols in Molecular Biology. p. 15.5.1-15.5.26. (1992).

Mullis K, Faloona F, Scharf S, Saiki R, Horn G, Erlich H: Specific enzymatic amplification of DNA *in vitro*: The polymerase chain reaction. Cold Spring Harbor Symposia on Quantitative Biology, volume LI. pp. 263-273. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, (1986).

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

Munns DN: Nodulation of *Medicago sativa* in solution culture II. Effects of nitrate on root hairs and infection. Plant Soil 29: 33 (1968).

Munthali M, Ford-Loyd BV, Newbury HJ: The random amplification of polymorphic DNA for fingerprinting plants. PCR Methods Appl 1: 274-276 (1992).

Murray et al: A comprehensive human linkage map with centimorgan density. Science 265: 2049-2054 (1994).

Myers RM, Maniatis T, Lerman LS: Detection and localization of single base changes by denaturing gradient gel electrophoresis. Methods-Enzymol 155: 501-527 (1987).

Nap J-P, Bisseling T: Nodulin function and nodulin gene regulation in root nodule development. In Gresshoff, PM (ed), Molecular Biology of Symbiotic Nitrogen Fixation. pp. 197-204. CRC Press Inc, Boca Raton, Forida (1990).

Nelson DL, Ledbetter SA, Corbo L, Victoria MM, Ramirez-Sole R, Webster TD, Ledbetter DH, Caskey CT: *Alu* polymerase chain reaction. A method for rapid isolation of human-specific sequences from complex DNA sources. Proc Natl Acad Sci USA 86: 6686-6690 (1989).

Nelson SF, McCusker JH, Sander MA, Kee Y, Modrich P, Brown PO: Genomic mismatch scanning. a new approach to genetic linkage mapping. Nat Genet 4: 11-18 (1993).

Nisson P, Rashtchian A, Watkins P: Rapid and efficient cloning of *Alu*-PCR products using uracil DNA glycosylase. PCR Meth Appl 1: 120-123 (1991).

O'Brien S. In O'Brien, S (ed), Genetic maps. Locus maps of complex genomes. Cold Spring Harbor Press, Cold Spring Harbor, USA (1990).

Olson M, Hood L, Cantor C, Botstein D: A common language for physical mapping of the human genome. Science 245: 1434-1435 (1989).

Olsson JE: Systematic control of soybean nodule autoregulation. PhD thesis Aust Nat Univ Canberra Australia (1988).

Olsson JE, Nakao P, Bohlool BB, Gresshoff PM: Lack of systematic suppression of nodulation in split root systems of supernodulating soybean (*G, max* [L.] Merr.) mutants. Plant Physiol 90: 1347-1352 (1989).

Orf JH, Mansur LM, Lark KG: Quantitative trait loci evaluation and identification of quantitative trait loci in a recombinant inbred soybean population. Abstract in the proceedings of the Molecular & Cellular Biology of the Soybean, 5th Biennial Conference, University of Georgia. p. 3. (1994).

Orita M, Suzuki Y, Sekiya T, Hayashi K: Rapid and sensitive detection of point mutations and DNA polymorphisms using the polymerase chain reaction. Genomics 5: 874-879 (1989).

Pääbo S, Irwin DM, Wilson AC: DNA damage promotes jumping between templates during enzymatic amplification. J Biol Chem 265: 4718-4721 (1990).

Palmer JD: Comparative organization of chloroplast genomes. Annu Rev Genet 19: 325-354 (1985).

Palmer RG, Hedges BR: Linkage map of soybean (*Glycine max* L. Merr.). In O'Brien, SJ (ed), Genetic maps: Locus maps of complex genomes. p. 6.139-6.148. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, USA (1993).

Paran I, Kesseli R, Michelmore R: Identification of restriction fragment length polymorphism and random amplified polymorphic DNA markers linked to downy mildew resistance genes in lettuce, using near-isogenic lines. Genome 34: 1021-1027 (1991).

Paran I, Michelmore RW: Development of reliable PCR-based markers linked to downy mildew resistance genes in lettuce. Theor Appl Genet 85: 985-993 (1993).

Parker CA: Evolution of nitrogen-fixing symbiosis in higher plants. Nature 179: 593-594 (1957).

Pé ME, Gianfranceschi L, Taramino G, Tarchini R, Angelini P, Dani M, Binelli G: Mapping quantitative trait loci (QTLs) for resistance to *Gibberella zeae* infection in maize. Mol Gen Genet 241: 11-16 (1993).

Penner GA, Bush A, Wise R, Kim W, Domier L, Kasha K, Laroche A, Scoles G, Molnar SJ, Fedak G: Reproducibility of random amplified polymorphic DNA (RAPD) analysis among laboratories. PCR Methods Appl 2: 341-345 (1993).

Perret X, Fellay R, Bjourson AJ, Cooper JE, Brenner S, Broughton WJ: Subtraction hybridization and shotgun sequencing: a new approach to identify symbiotic loci. Nucl Acids Res 22: 1335-1341 (1994).

Petit A, Stougaard J, Kühle A, Marckee KA, Tempe J: Transformation and regeneration of the legume *Lotus corniculatus*: A system for molecular analysis of symbiotic nitrogen fixation. Mol Gen Genet 207: 245-250 (1987).

Pierce M, Bauer, WD: A rapid regulatory response governing nodulation in soybean. Plant Physiol 73: 286-290 (1983).

Postgate JR: The nitrogen-fixing bacteria. In Postgate, J (ed), The Fundamentals of Nitrogen Fixation. pp. 5-19. Cambridge University Press, Cambridge, UK (1982).

Postma JG, Jacobsen E, Feenstra WJ: Three pea mutants with an altered nodulation studied by genetic analysis and grafting. J Plant Physiol 132: 424 (1988).

Prabhu RR, Gresshoff PM: Inheritance of polymorphic markers generated by DNA amplification fingerprinting and their use as genetic markers in soybean. Pl Mol Biol 26: 105-116 (1994).

Pracht JE, Nickell CD, Harper JE: Genes controlling nodulation in soybean: Rj_5 and Rj_6. Crop Sci 33: 711-713 (1993).

Putterill J, Robson F, Lee K, Coupland G: Chromosome walking with YAC clones in *Arabidopsis*: isolation of 1700 kb of contiguous DNA on chromosome 5, including a 300 kb region containing the flowering-time gene *CO*. Mol Gen Genet 239: 145-157 (1993).

Qiu J, Weaver DB, Tuzun S: Identification of DNA amplification fingerprinting markers linked to resistance to frog-eye leaf spot in soybean. Abstract in the proceedings of the Molecular & Cellular Biology of the Soybean, 5th Biennial Conference, University of Georgia. p. 40. (1994).

Rafalski JA, Tingey SV: Genetic diagnosis in plant breeding: RAPDs, microsatellites and machines. Trends Gen 9: 275-280 (1993).

Ragab RA, Dronavalli S, Saghai-Maroof MA, Yu YG: Construction of a sorghum RFLP linkage map using sorghum and maize DNA probes. Genetics 37: 590-594 (1994).

Ralston EJ, English JJ, Dooner HK: Sequence of three bronze alleles of maize and correlation with the genetic fine structure. Genetics 119: 185-197 (1988).

Reiter RS, Williams JGK, Feidmann KA, Rafalski A, Tingey S, Scolnik VS: Global and local genome mapping in *Arabidopsis thaliana* by using recombinant inbreds and random amplified polymorphic DNAs. Proc Natl Acad Sci 89: (1992).

Richards EJ, Ausubel FM: Isolation of a higher eukaryotic telomere from *Arabidopsis thaliana*. Cell 53: 127-136 (1988).

Richards RI, Sutherland GR: Dynamic mutations: a new class of mutations causing human disease. Cell 70: 709-712 (1992).

Ridge R, Rolfe BG: Sequence of events during the infection of the tropical legume *Macroptilium atropurpureum* Urb. by broad-host-range, fast growing *Rhizobium ANU240*. J Plant Physiol 122: 121-137 (1985).

Riedy MF, Hamilton J, Aquadro CF: Excess of non-parental bands in offspring from known primate assayed using RAPD PCR. Nucleic Acids Res 4: 918 (1992).

Rieseberg LH, Hichang C, Chan R, Spore C: Genomic map of a diploid hybrid species. Heredity 70: 285-293 (1993).

Riordan JR, Rommens JM, Kerem B-S, Alon N, Rozmahel R, Grzelezak Z, Zielenski J, Lok S, Plavsic N, Chou J-L, Drumm ML, Iannuzzi MC, Collins FS, Tsui L-C: Identification of the cystic fibrosis gene: cloning and characterization of complimentary DNA. Science 245: 1066-1075 (1989).

Rivard SR, Cappadocia M, Vincent G, Brisson N, Landry BS: Restriction fragment length polymorphism (RFLP) analysis of plants produced by *in vitro* culture of *Solanum chacoense* Bitt. Theor Appl Genet 78: 49-56 (1989).

Robertson DS: A possible technique for isolation genomic DNA for quantitative genetic traits. Theor Biol 117: 1-3 (1985).

Rolfe BG, Gresshoff PM: Genetic analysis of legume nodule initiation. Ann Rev Plant Physiol-Plant Mol Biol 39: 297-319 (1988).

Rommens JM, Iannuzi MC, Kerem B-S, Drumm ML, Melmer G, Dean M, Rozmahel R, Cole JL, Kennedy D, Hidaka N, Zsiga M, Buchwald M, Riordan JR, Tsui L-C, Collins FS: Identification of the cystic fibrosis gene: chromosome walking and jumping. Science 245: 1059-1065 (1989).

Ronald P: Perspectives of map-based cloning of rice genes. DNA Link 3: 35-39 (1993).

Ronald PC, Albano B, Tabien R, Abenes L, Wu KS, McCouch S, Tanksley SD: Genetic and physical mapping of the rice bacterial blight resistance locus *Xa21*. Mol Gen Genet 236: 113-120 (1993).

Roth EJ, Frazier BL, Apuya NR, Lark KG: Genetic variation in an inbred plant: variation in tissue cultures of soybean [*Glycine max* (L.) Merrill]. Genetics 121: 359-368 (1989).

Roy A, Frascaria N, MacKay J, Bousquet J: Segregating random amplified polymorphic DNAs(RAPDs) in *Betula alleghaniensis*. Theor Appl Genet 85: 173-180 (1992).

Saghai-Maroof MA, Biyashev RM, Yang GP, Zhang Q, Allard RW: Extraordinarily polymorphic microsatellite DNA in barley: species diversity, chromosomal location and population dynamics. Proc Natl Acad Sci USA 91: 5466-5470 (1994).

Saiki RK, Scharf S, Falloona F, Mullis KB, Horn G, Erlich HA, Arnheim N: Enzymatic amplification of ß-globin genomic sequences and restriction site analysis for diognosis of sickle cell anemia. Science 230: 1350-1356 (1985).

Sambrook J, Fritsch EF, Maniatis T. Molecular Cloning: A Laboratory Manual Cold Spring Harbor Laboratory, Cold Spring Harbor, NY (1989).

Sánchez F, Padilla JE, Pérez H, Lara M: Control of nodulin genes in root-nodule development and metabolism. Ann Rev Plant Physiol-Plant Mol Biol 42: 507-528 (1991).

Sargent I, Huang SZ, Rolfe BG, Djordjevic MA: Split-root assays using *Trifolium subterraneum* show that *Rhizobium* infection induces a systemic response that can inhibit nodulation of another invasive*Rhizobium* strain. Appl Environ Micro 53: 1611-1619 (1987).

Sayavedra-Soto LA, Angermüller SA, Prabhu RR, Gresshoff PM: Polypeptide patterns in leaves of *Glycine max* (L.) Merr. cv Bragg and its supernodulating mutant during early nodulation. Physiol Mol Biol Plants 1: 27-36 (1995).

Schachermayr G, Siedler H, Gale MD, Winzler H, Winzler M, Keller B: Identification and localization of molecular markers linked to the *Lf9* leaf rust resistance gene of wheat. Theor Appl Genet 88: 110-115 (1994).

Scheres B, Weil C, Zalenski A, Horvath B, Spaink H: The *ENOD12* product is involved in the infection process during the pea-*Rhizobium* interaction. Cell 60: 281-294 (1990).

Schwartz DC, Cantor CR: Separation of yeast chromosome sized DNAs by pulsed-field gradient gel eletrophoresis. Cell 37: 76-77 (1984).

Scott MP, Haymes KM, Williams SM: Parentage analysis using RAPD PCR. Nucl Acids Res 20: 5493 (1992).

Sherwood JE, Truchet GL, Dazzo FB: Effect of nitrate supply on the in vivo synthesis and distribution of trifolin A, a *Rhizobium trifolii*-binding lectin, in *Trifolium repens* seedlings. Planta 162: 540 (1984).

Shi L, Zhu T, Keim P: Painting soybean chromosomes by FISH with high-copy DNA probes. Abstract in the proceedings of the Molecular & Cellular Biology of the Soybean, 5th Biennial Conference, University of Georgia. p. 1. (1994).

Shizuya H, Birren B, Kim VJ, Mancino V, Slepak T, Tachiri Y, Simon M: Cloning a stable maintenance of 300 kilobase-pair fragment of human DNA in *Escherichia coli* using an F-factor based vector. Proc Natl Acad Sci USA 89: 8794-8797 (1992).

Shoemaker RC, Guffy RD, Lorenzen L, Specht JE: Molecular genetic mapping of soybean: Map utilization. Crop Sci 32: 1091-1098 (1992).

Shoemaker RC, Olson TC: Molecular linkage map of soybean (*Glycine max* L. Merr.). In O'Brien, SJ (ed), Genetic maps: Locus maps of complex genomes. p. 6.131-6.138. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, USA (1993).

Shoemaker RC, Specht JE: Integration of the soybean molecular and classical geentic linkage groups. Crop Sci In press (1994).

Singh RJ, Hymowitz T: The genomic relationship between *Glycine max* (L) Merr. and *G. soja* Sieb. and Zucc. as revealed by pachytene chromosome analysis. Theor Appl Genet 76: 705-711 (1988).

Singleton PW, Kessel van C: Effect of localized nitrogen availability to soybean half-root systems on photosynthate partitioning to roots and nodules. Plant Physiol 83: 552-556 (1987).

Skorupska HT, Mahalingam R, Choi IS, Warner A: RAPD markers for resistance to soybean cyst nematode *Heterodera glycines* I. and marker assisted selection. Abstract in the proceedings of the Molecular & Cellular Biology of the Soybean, 5th Biennial Conference, University of Georgia. p. 27. (1994).

Skroch P, Tivang J, Niehuis J: Analysis of genetic relationships using RAPD marker data. In Neff, M (ed), Application of RAPD Technology in Plant Breeding. pp. 26-30. ASHS Publishers, St Paul, MN (1993).

Slocum MK, Figdore SS, Kennard WC, Suzuki JY, Osborn TC: Linkage arrangement of restriction fragment length polymorphism loci in *Brassica oleracea*. Theor Appl Genet 80: 57-64 (1990).

Smit G, Van Brussel T, Kijne JW: Inactivation of a root factor by ineffective *Rhizobium*: A molecular key to autoregulation of nodulation in *Pisium sativum*. In Palacios, R, Mora, J, Newton, WE (eds), New Horizons in Nitrogen Fixation. Current Plant Science and Biotechnology in Agriculture. Kluwer Publishing, Netherlands (1993).

Smith S, Chin E: The utility of random primer-mediated profiles, RFLPs and other technologies to provide useful data for varietal protection. In Neff, M (ed), Application of RAPD technology in plant breeding. pp. 46-49. ASHS Publishers, St Paul, MN (1992).

Soller M, Beckmann JS: Genetic polymorphism in varietal identification and genetic improvement. Theor Appl Genet 67: 25-33 (1983).

Sommer R, Tautz D: Minimal homology requirements for PCR primers. Nucl Acids Res 17: 6749 (1989).

Strege M, Lagu A: Separation of DNA restriction fragments by capillary electrophoresis using coated fused silica capillaries. Anal Chem 63: 1233-1236 (1991).

Suzuki H, Verma DPS: Soybean nodule-specific uricase (nodule-35) is expressed and assembled into a functional tetrameric holoenzyme in *Escherichia coli*. Plant Physiol 95: 384-389 (1991).

Tang X, Tashiro H, Eki T, Murakami Y, Soeda E, Sakakura T, Watkins PC, Yokoyama K: Generation of 19 STS markers that can be anchored at specific sites on chromosome 21. Genomics 14: 185-187 (1992).

Tanksley SD: High-density molecular linkage maps of the tomato and potato genomes. Genetics 132: 1141-1160 (1992).

Tanksley SD, Causse M, Fulton T, Ahn N, Wang Z, Wuk A, Xiao J, Seconda G, McCouch S: A high density molecular map of the rice genome. Rice Genet 9: 111-115 (1992).

Tanksley SD, Ganal MW, Martin GB: Chromosome landing: a paradigm for map-based gene cloning in plants with large genomes. Trends Genet 11: 64-69 (1995).

Thormann CE, Ferreira ME, Camargo LEA, Tivang JG, Osborn TC: Comparison of RFLP and RAPD markers for estimating gentic relationships within and among Cruciferous species. Theor Appl Genet 88: 973-980 (1994).

Torrey JG: Endophyte sporulation in root nodules of actinorrhizal plants. Physiol Plant 70: 279-283 (1987).

Tragoonrung S, Kanazin V, Hayes PM, Blake TK: Sequence-tagged-site-facilitated PCR for barley genome mapping. Theor Appl Genet 84: 1002-1008 (1992).

Trinick MJ: Structure of nitrogen-fixing nodules formed by *Rhizobium* on roots of *Parasponia andersonii* Planch. Can J Microbiol 25: 565-578 (1979).

Truchet G, Barker DG, Camut S, deBilly F, Vasse J, Hugnet T: Alfalfa nodulation in the absence of *Rhizobium*. Mol Gen Genet 219: 65-68 (1989).

Tulseiram LK, Glaubitz JC, Kiss G, Carlson JE: Single tree genetic linkage mapping in conifers using haploid DNA from megagametophytes. Bio/technology 10: 686-689 (1992).

Tulsieram L, Compton WA, Morris R, Thomas-Compton M, Eskridge K: Analysis of genetic recombination in maize populations using molecular markers. Theor Appl Genet 82: 65-72 (1991).

Uphoff H, Wricke G: Random amplified polymorphic DNA (RAPD) markers in suger beet (*Beta vulgaris* L.): mapping the genes for nematode resistance and hypocotyl color. Plant Breed 109: 168-171 (1992).

Vallejos CE, Sakiyama NS, Chase CD: A molecular marker-based linkage map of *Phaseolus vulgaris* L. Genetics 131: 733-740 (1992).

Vest G: Rj_3-A gene conditioning ineffective nodulation in soybean. Crop Sci 10: 34-35 (1970).

Vest G, Caldwell BE: Rj_4-A gene conditioning ineffective nodulation in soybean. Crop Sci 12: 692-693 (1972).

Walden R, Hayashi H, Schell J: T-DNA as a gene tag. Plant J 1: 281-288 (1991).

Wang G-L, Holstein TE, Song W-Y, Wang H-P, Ronald PC: Construction of a rice bacterial artificial chromosome library and identification of clones linked to the *Xa-21* disease resistance locus. Plant J 7: 525-533 (1995).

Wang Z, Weber JL, Zhong G, Tanksley SD: Survey of plant short tandem DNA repeats. Theor Appl Genet 88: 1-6 (1994).

Watanabe I, Espinas CR, Berja NS, Alimagno BV: Utilization of the *Azolla-Anabena* complex as a nitrogen fertilizer for rice. IRRI Research Paper series #11. International Rice Research Institute, Manila, Phillipines (1977).

Weaver KR: DNA amplification fingerprinting of centipedegrass. MS Thesis, University of Tennessee (1993).

Weaver KR, Caetano-Anollés G, Gresshoff PM, Callahan LM: Isolation and cloning of DNA amplification products from silver stained polyacrylamide gels. Biotechniques 16: 226-227 (1994).

Weeden NF, Timmerman GM, Hemmat M, Kneew BE, Lodhi MA: Inheritance and reliablity of RAPD markers. In Neff, M (ed), Application of RAPD Technology in Plant Breeding. pp. 26-30. ASHS Publishers, St Paul, MN (1992).

Weising K, Nybom H, Wolff K, Meyer W. DNA Fingerprinting in Plants and Fungi. CRC Press, Boca Raton, Florida, USA (1995).

Welsh J, McClelland M: Fingerprinting genomes using PCR with arbitrary primers. Nucl Acids Res 18: 7213-7218 (1990).

Welsh J, McClelland M: Genomic fingerprints using arbitrarily primed PCR and a matrix of pairwise combination of primers. Nucl Acids Res 19: 5725-5729 (1991).

Welsh J, Honeycutt M, McClelland M, Sobral BWS: Parentage determination in maize hybrids using the arbitrarily primed polymerase chain reaction (AP-PCR). Theor Appl Genet 82: 473-476 (1991).

Whitkus R, Doebley J, Lee M: Comparative mapping of sorghum and maize. Genetics 132: 1119-1130 (1992).

Williams CE, St Clair DA: Phenetic relationships and levels of variability detected by restriction fragment length polymorphism and random amplified polymorphic DNA analysis of cultivated and wild accessions of *Lycopersicon esculentum*. Genome 36: 619-630 (1993).

Williams JGK, Kubelik AR, Lavak KJ, Rafalski JA, Tingey SY: DNA polymorphisms amplified by arbitrary primers are useful as genetic markers. Nucl Acids Res 18: 6531-6535 (1990).

Williams JGK, Hanafey MK, Rafalski JA, Tingey SV: Genetic analysis using RAPD markers. In Wu, R (ed), Methods in Enzymology, Recombinant DNA Volume, vol 218. pp. 704-740. Academic Press, Inc, San Diego, CA, USA (1991).

Williams JGK, Reiter RS, Young RM, Scolnik PA: Genetic mapping of mutations using phenotypic pools and mapped RAPD markers. Nucl Acids Res 21: 2697-2702 (1993).

Williams LF, Lynch DL: Inheritance of a non-nodulating character in the soybean. Agron J 46: 28-29 (1954).

Williams MNV, Pande N, Nair S, Mohan M, Bennet J: Restriction fragment length polymorphism analysis of polymerase chain reaction products amplified from mapped loci of rice (*Oryza sativa* L.) genomic DNA. Theor Appl Genet 82: 489-498 (1991).

Wooster R, Neuhausen SL, Mangion J, Quirk Y, Ford D, Collins N, Nguyen K, Seal S, Tran T, Averill D et al: Localization of a breast cancer susceptibility gene *BRCA2* to chromosome 13q12-13. Science 265: 2088-2090 (1994).

Wu DY, Ugozzoli L, Pal BK, Wallace RB: Allele-specific enzymatic amplification of beta-globin genomic DNA for diagnosis of sickle cell anemia. Proc Natl Acad Sci USA 86: 2757-2760 (1989).

Xu SJ, Singh RJ, Hymowitz T: Research on primary trisomics in soybean. Abstract in the proceedings of the Molecular & Cellular Biology of the Soybean, 5th Biennial Conference, University of Georgia. p. 16. (1994).

Yadava JS, Chowdhury JB, Kakar SN, Nainawatece HS: Comparative electrophoretic studies of proteins and enzymes of some *Brassica* species. Theor Appl Genet 54: 89-91 (1979).

Yang WC, Katinakis P, Hendriks P, Smolders A, de Vries F, Spee J, van Kammen A, Bisseling T, Franssen H: Characterization of *GmENOD40*, a gene showing novel patterns of cell-specific expression during soybean development. Plant J 3: 573-585 (1993).

Yang X, Quiros C: Identification of celery cultivars with RAPD markers. Theor Appl Genet 86: 205-212 (1993).

Yanofski MF, Ma H, Bowman JL, Drews GN, Feldman KA, Meyerowitz EM: The protein encoded by *Arabidopsis* homeotic gene *agamous* resembles transcription factors. Nature 346: 35-39 (1990).

Yu K, Pauls KP: Identification of a RAPD marker associated with somatic embryogenesis in alfalfa. Pl Mol Biol 22: 269-277 (1993).

Yu K, Pauls KP: Rapid estimation of genetic relatedness among heterogenous populations of alfalfa by random amplification of bulked genomic DNA samples. Theor Appl Genet 86: 788-794 (1993).

Zabeau M: Selective restriction fragment amplification: a general method for DNA fingerprinting. European Patent Application Publication No 0 534 858A1 (1993).

Zamir D and Tanksley SD: Tomato genome is comprised largely of fast evolving low copy-number sequences. Mol Gen Genet 213: 254-261 (1988).

Zhao X, Kochert G: Characterization and genetic mapping of a short, highly-repeated, interspersed DNA sequence from rice (*Oryza sativa* L.). Mol Gen Genet 231: 353-359 (1992).

Zimmermann JL, Goldberg RB: DNA sequence organization in the genome of *Nicotiana tabacum*. Chromosoma 59: 227-231 (1977).

VITA

Rekha Prabhu was born in Bombay, India and graduated from St. Agnes' High School in Bombay in 1973. She obtained a bachelor's degree from St. Xavier's College, Bombay with a major in Chemistry and a minor in Microbiology in 1977. She later obtained a Master's degree in Biochemistry from the University of Bombay in 1980. Rekha came to Knoxville in Fall 1982. She obtained a Master's degree in Biochemistry in December 1985 from the University of Tennessee. Her Master's thesis was entitled "Direct measurement of the energetics of association between myelin basic protein and phosphatidylserine vescicles". During this period she taught undergraduate and graduate courses as a teaching assistant. Rekha joined Dr. Gresshoff's laboratory in Fall 1989. The doctoral degree was received in August 1995.