ISOLATION OF THE NUCLEOTIDE-BINDING SITE FAMILY OF RESISTANCE GENE ANALOGS FROM ALFALFA

by

JUAN CARLOS CORDERO ARTIÑANO

B.S., Kansas State University, 1997

A THESIS

Submitted in partial fulfillment of the
requirements for the degree

MASTER OF SCIENCE

Department of Genetics
College of Agriculture

KANSAS STATE UNIVERSITY
Manhattan, Kansas

2000

Approved by:

Co-major Professor
Daniel Z. Skinner

Approved by:

Co-major Professor
George H. Liang

nich univ
LD2668
.T4
AGRN
2000
C67

ABSTRACT

Research findings over the past decade revealed exciting information regarding disease and pest resistance in plants. Scientists found that homology exists between genes (R genes) of different plant species that confer strain-specific resistance against a wide range of pathogens and pests including: viruses, bacteria, fungi, nematodes and insects. These genes have been found to be a strong, naturally occurring source of resistance in plants with great potential for environmentally safe applications in crop protection.

The purpose of this study was to investigate the existence and expression of R genes in alfalfa (*Medicago sativa* L.) and to analyze the diversity in alfalfa and in the genus *Medicago.* We used a PCR approach with degenerate primers designed from the highly conserved motifs, P-Loop and GLPL, found within the nucleotide binding site (NBS) of NBS-LRR resistance genes. The degenerate PCR approach was used successfully on genomic DNA and cDNA from *Medicago sativa* and genomic DNA from thirteen other *Medicago* species. Sequence data from ninety-four clones of the NBS PCR product provided 51 *Medicago sativa* resistance gene analogs (MSRGA's) sharing from 35% to 99% homology. The eighteen families that share less than 95% homology represent the 51 MSRGA's. Additionally, dendrogram and bootstrap analysis showed that the 51 MSRGA's grouped in six major clusters that suggested three different evolutionary origins. Transcription of three of the MSRGA families was confirmed by RT-PCR. Diversity in resistance gene analogs between species of the genus *Medicago* was confirmed by restriction pattern analysis of the degenerate PCR products.

The characterization of the MSRGA family in alfalfa and the diversity of resistance gene analogs in the genus *Medicago* provided by this study represent a pool of candidate resistance genes which will be fundamental for the development of disease and pest resistant alfalfa cultivars in the future.

TABLE OF CONTENTS

LIST OF FIGURES..... iii

LIST OF TABLES..... v

ACKNOWLEGEMENTS..... vi

DEDICATION..... vii

INTRODUCTION..... 1

IMPACT OF DISEASES AND PESTS ON ALFALFA AND OTHER MAJOR CROPS..... 3

GENETICS OF DISEASE AND PEST RESISTANCE IN PLANTS..... 6

MOLECULAR GENETICS OF PLANT-PATHOGEN INTERACTIONS..... 8

DISEASE RESISTANCE GENES..... 10

DOMAINS OF RESISTANCE GENES ENCODING A NUCLEOTIDE BINDING SITE..... 13

TOLL-INTERLEUKIN RECEPTOR..... 13

LEUZINE ZIPPER..... 13

LEUCINE-RICH REPEAT..... 14

NUCLEOTIDE-BINDING SITE..... 15

CANDIDATE DISEASE RESISTANCE GENES..... 19

MATERIALS AND METHODS..... 21

ISOLATION OF *MEDICAGO SATIVA* RESISTANCE GENE ANALOGS..... 21

ISOLATION OF EXPRESSED *MEDICAGO SATIVA* RESISTANCE GENE ANALOGS..... 24

RESISTANCE GENE ANALOG DIVERSITY WITHIN THE GENUS *MEDICAGO*..... 26

RESULTS AND DISCUSSION..... 27

MEDICAGO SATIVA RESISTANCE GENE ANALOG NUCLEOTIDE SEQUENCES..... 31

MEDICAGO SATIVA RESISTANCE GENE ANALOG ENCODED SEQUENCES..... 47

MEDICAGO SATIVA RESISTANCE GENE ANALOG AND THE NUCLEOTIDE-BINDING SITE OF CHARACTERIZED RESISTANCE GENES..... 54

ISOLATION OF EXPRESSED *MEDICAGO SATIVA* RESISTANCE GENE ANALOGS..... 61

RESISTANCE GENE ANALOG DIVERSITY WITHIN THE GENUS *MEDICAGO*..... 65

CONCLUSIONS..... 68

REFERENCES..... 74

LIST OF FIGURES

Figure 1 Pathogens, pests and diseases in plants. 4

Figure 2 Plant-pathogen molecular interactions and the plant's defense response mechanism. 8

Figure 3 Classification of plant disease resistance genes according to the encoded protein structure. 10

Figure 4 Nucleotide-binding site of disease resistance genes. 16

Figure 5 Nucleotide-binding site amplification by degenerate PCR on alfalfa DNA. 27

Figure 6 NBS degenerate PCR product digested with Tsp 509I. 28

Figure 7 NBS clones amplified by PCR with M13 primers. 29

Figure 8 Individual NBS clones digested with Tsp509I. 30

Figure 9a Clustal W alignment of bases 1-50 of MSRGA's. 35

Figure 9b Clustal W alignment of bases 51-97 of MSRGA's. 36

Figure 9c Clustal W alignment of bases 98-145 of MSRGA's. 37

Figure 9d Clustal W alignment of bases 146-195 of MSRGA's. 38

Figure 9e Clustal W alignment of bases 196-244 of MSRGA's. 39

Figure 9f Clustal W alignment of bases 245-293 of MSRGA's. 40

Figure 9g Clustal W alignment of bases 294-343 of MSRGA's. 41

Figure 9h Clustal W alignment of bases 344-393 of MSRGA's. 42

Figure 9i Clustal W alignment of bases 394-443 of MSRGA's. 43

Figure 9j Clustal W alignment of bases 444-475 of MSRGA's. 44

Figure 10 Dendrogram generated from alignment of nucleotide sequences of MSRGA's. 45

Figure 11a Clustal W alignment of residues 1-44 of MSRGA encoded sequences. 50

Figure 11b Clustal W alignment of residues 45-90 of MSRGA encoded sequences. 51

Figure 11c Clustal W alignment of residues 91-140 of MSRGA encoded sequences. 52

Figure 11d Clustal W alignment of residues 141-158 of MSRGA encdoded sequences. 53

Figure 12 Dendrogram generated from alignment of MSRGA encoded sequences. 59

Figure 13 NBS amplification from alfalfa first strand cDNA. 61

Figure 14 Individual NBS cDNA clones amplified with M13 primers. 62

Figure 15 Individual NBS cDNA clones digested with Tsp509I. 63

Figure 16 NBS amplification by degenerate PCR on species belonging to the genus *Medicago*. 65

Figure 17 NBS degenerate PCR product from *Medicago* species digested with Tsp509I. 66

LIST OF TABLES

Table1	Genetics of host-pathogen interactions in the gene-for-gene model of resistance.	7
Table 2	Classification of the most studied disease and pest resistance genes. .	12
Table 3	Characteristic motifs of the nucleotide-binding site in Plants.	18
Table 4	GenBank Accession numbers of resistance gene analogs (RGA) from *Medicago sativa.*	23
Table 5	Classification of the MSRGA clones from Medicago sativa.	34
Table 6	Nucleotide homology between MSRGA families.	46
Table 7	Amino acid homology between MSRGA's and R genes from diverse plant species.	60

ACKNOWLEDGMENTS

The five years spent at Kansas State University have been an unforgettable experience from which I will always value the knowledge and cherish the friendships made. For this, I would like to thank all the people with whom I've shared moments at Kansas State University.

I would like to thank the members of my committee Dr. Scot H. Hulbert and Dr. George H. Liang for all their time, guidance and support. I would also like to thank all my professors from both the Agronomy and Plant Pathology departments for their commitment towards the formation and guidance of us students.

A very special thank you to my major professor Dr. Daniel Z. Skinner. Your guidance, encouragement and friendship have made my studies at Kansas State University an outstanding experience that will always be appreciated.

I would also like to thank all my colleagues and friends for always being available for help, and encouragement. Your friendship during my years at KSU made the years here more fun and enjoyable.

Finally, I would like to thank my family and friends at Costa Rica for their love and support and for always being there regardless of time or place.

DEDICATION

To my family, for all your love, support and encouragement.

INTRODUCTION

Alfalfa, *Medicago sativa*, has been cultivated by man before recorded history. The oldest records of alfalfa, from Turkey, date back to 1300B.C., and it is believed to have been cultivated long before that. Alfalfa was valued as being a highly nutritious feed for animals. Trade between civilizations spread the valued alfalfa from its place of origin near Iran to distant geographic locations throughout the world.

In the present, alfalfa is grown for forage in most temperate countries around the world. In countries where it is well adapted, alfalfa is recognized as the most important forage crop species due to its high protein values, nitrogen fixing abilities, winter hardiness, high yields, protein yields per area, and as a source of nectar for honey production.

In the USA, alfalfa now is the most widely grown forage crop. It is the fourth most grown crop in terms of area under production and total production, being surpassed only by soybeans, wheat, and corn (NASS, 1999). For these reasons alfalfa has raised interest among producers, scientists, government agencies and industry, all of which have contributed to the improvement of the "Queen of forages."

Efforts in the improvement of alfalfa began soon after it was first introduced in the United States in the 19th century. After its first introduction, plant explorers collected additional sources of germplasm from around the world.

This wide genetic pool provided breeders with valuable genetic variation that laid the foundation for alfalfa breeding in the USA. Breeders have achieved great progress in yields obtained from alfalfa throughout this century. Nevertheless, the achievements made are much lower than those achieved for the other major crops in the USA, such as

corn and wheat. Most of the progress made in alfalfa has been in adaptation of the crop to different environments, tolerance to frequent harvests, and resistance to diseases and pests; which has increased yields by reducing losses, but not by increasing yield *per se* (Hill et al. 1988). It is difficult to transmit high yielding qualities to the new cultivars, if the major focus in the breeding program is to make the crop resistant to a severe pathogen or pest. Improvement of alfalfa in the past has been based on traditional breeding methods. Advances in plant breeding, molecular biology, plant transformation, and recent findings in molecular interactions of host and pathogens will optimize the transfer of disease and pest resistance to desirable varieties that do not perform well due to disease or pest susceptibility.

IMPACT OF DISEASES AND PESTS ON ALFALFA AND OTHER MAJOR CROPS

Ever since man relied on crop production as its major food source, we have had to cope with the adverse effects of pathogens and pests on plant growth. Illustrations of this antagonistic relationship are found in ancient records such as the Bible but have been dealt with since the first civilizations appeared that relied heavily on agriculture.

Today, thousands of years have passed and man still relies heavily on agriculture and still has to face the factors that affect negatively the efficiency of plant production. The importance of agriculture has spanned beyond food production and now reaches other sectors as textiles, wood, paper, pharmaceuticals, and other derivatives. Little has changed however, in the relationship that we have with plants and pathogens.

Lack of genetic diversity in our crops as well as the practice of monoculture has increased the vulnerability of agriculture to pests and diseases. Pressure on increasing world agricultural production brought great advances in genetics and breeding which produced superior cultivars for each crop. Nevertheless, each time those superior cultivars become available they are widely grown making the genetic makeup of our crops extremely uniform. New, improved cultivars become available on a regular basis but are then widely grown, again leading to genetic uniformity. This way the genetic foundation becomes narrower as pressure on breeding new cultivars becomes more intense. The genetic uniformity creates a risky situation that enables new pathogens to colonize plantations causing significant losses to the new cultivars.

Alfalfa, like every other crop, constantly is challenged by pathogens and pests that limit the efficiency of alfalfa production (Figure1). In 1988, 37 diseases, 5 nematodes,

and more than a hundred insects (Manglitz and Ratcliffe 1988) were known that cause damage to alfalfa. Of the 37 alfalfa diseases, 28 were caused by fungi, 3 by bacteria, 5 by viruses, and one by mycoplasma (Leath et al. 1988).

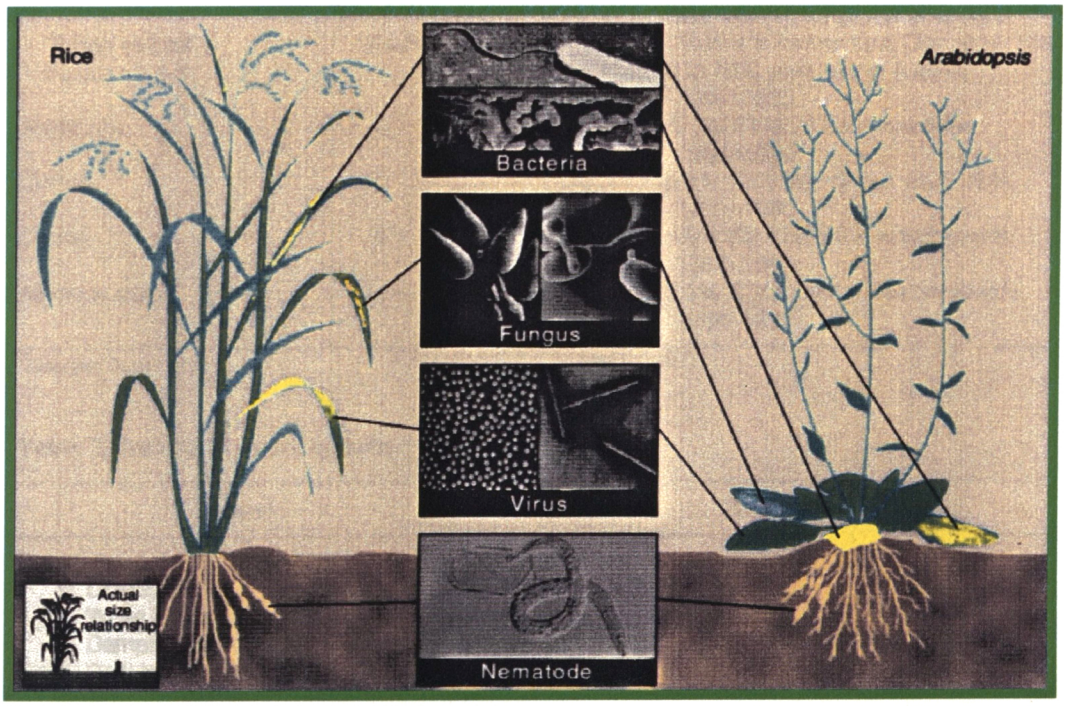


Figure 1: Bacteria, fungi, viruses, nematodes and insects attack our food crops, producing different symptoms and diseases which translate into economic losses to producers. Despite the large differences between plant pathogens, scientists have found that there are similarities in the plant's defense response against them. Reprinted with permission from Baker et al. 1997. Copyright 1997 American Association for the Advancement of Science.

As mentioned earlier, a major portion of the improvement of alfalfa has come from the development of pest and disease resistant cultivars. Nevertheless, alfalfa production still faces considerable losses from diseases and pests. It has been estimated that over 25% of the yearly production of alfalfa hay and 10% of the annual seed crop is lost to diseases (Leath et al. 1988). In 1972, it was estimated that insect damage represented a cost of no less than $260 million (Manglitz and Ratcliffe 1988).

These estimates indicate that there is still a lot of ground for improving alfalfa production in the USA and other countries that rely on the cultivars developed in the

USA. The fact that alfalfa is not considered a high value crop makes the use of chemical control against diseases economically unfeasible. Furthermore, there is growing concern in the general public of the long-term effects agricultural chemicals might cause to human health and the environment.

For these reasons, development of pest and disease resistant cultivars still remains the most economical and safest form of disease and pest prevention. The complex genetic characteristics in alfalfa provide many obstacles for the completion of this task. However, in the past ten years, scientists have discovered exciting information concerning the molecular genetics of plant and pathogen interactions. A good understanding of the recent findings of plant pathogen interactions will be fundamental for the improvement of alfalfa, and more specifically in the development of disease and pest resistant cultivars.

GENETICS OF DISEASE AND PEST RESISTANCE IN PLANTS

The advances in genetics by Gregor Johann Mendel in the 19th century helped man understand the scientific basis of heredity, which helped plant breeders reach unforeseeable yield levels in agriculture. Since early in the 20th century plant breeders recognized that a plant's resistance to diseases is inherited from parents to progeny in a Mendelian fashion.

In the 1940's a plant breeder named H.H. Flor, working with flax, *Linum usitatsissimum*, and its fungal pathogen, *Melampsora lini,* studied host and pathogen race differentials. Flor noted that for every gene in the plant race differential that conferred resistance to a specific strain of pathogen, there was a gene in the pathogen that conferred avirulence. This concept has been known as the gene-for-gene model of resistance (Flor 1956).

The genes responsible for providing avirulence in a pathogen are now known as avirulence genes. The term avirulence means the lack of virulence to a plant. In other words, the avr genes make the pathogen avirulent in the presence of its counterpart, an R gene, which is in charge of conferring resistance in the plant. This genetic relationship which constitutes the basis for the gene-for-gene concept can be summarized with four different scenarios possible (Table 1). It is important to point out that most frequently avr and R genes have been demonstrated to be dominant alleles. First, if the pathogen has a functional avirulence gene, avr+, and attacks a plant that has a functional resistance gene R+, a defense response is elicited in the plant. This is known as a resistant or incompatible interaction between host and pathogen. Second, if the same pathogen with avr+ attacks a plant that lacks the resistance gene, R-, then disease develops, and we have

a susceptible or compatible interaction between both parts. Third, if the pathogen that lacks the avirulence gene, avr-, attacks a plant that has the resistance gene, R+, then disease develops giving a susceptible or compatible interaction. In this case, even though the plant possesses the corresponding R gene it is unable to recognize the pathogen and elicit a defense response, because the pathogen lacks the avr gene. Fourth, if a pathogen that lacks the avirulence gene, avr-, attacks a plant that lacks the resistance gene, R-, then disease develops, giving a susceptible or compatible interaction. This concept is summarized in the table below.

Table 1: Genetics of host-pathogen interactions in the gene-for-gene model of resistance.

Genotype	**Pathogen AVR+**	**Pathogen avr-**
Plant R+	Resistant	Susceptible
Plant r-	Susceptible	Susceptible

This genetic relationship gets more complex when more than one R gene and avr gene are involved; however, the basic principle as explained previously still holds. An important consideration to keep in mind is that no matter how many avr-R gene pairs, one pair suffices to elicit the defense response in the plant resulting in a resistant phenotype.

MOLECULAR GENETICS OF PLANT PATHOGEN-HOST INTERACTIONS

Flor's concept explained the existence of specific resistance and avirulence genes. The gene-for-gene model is now widely accepted and the first genes that fit Flor's model have been isolated and characterized at the molecular level.

Flor's model, together with findings from later scientists who have deciphered much of the gene-for-gene model at the molecular level, have helped us to understand the plant's defense mechanism upon pathogen attack (Figure 2).

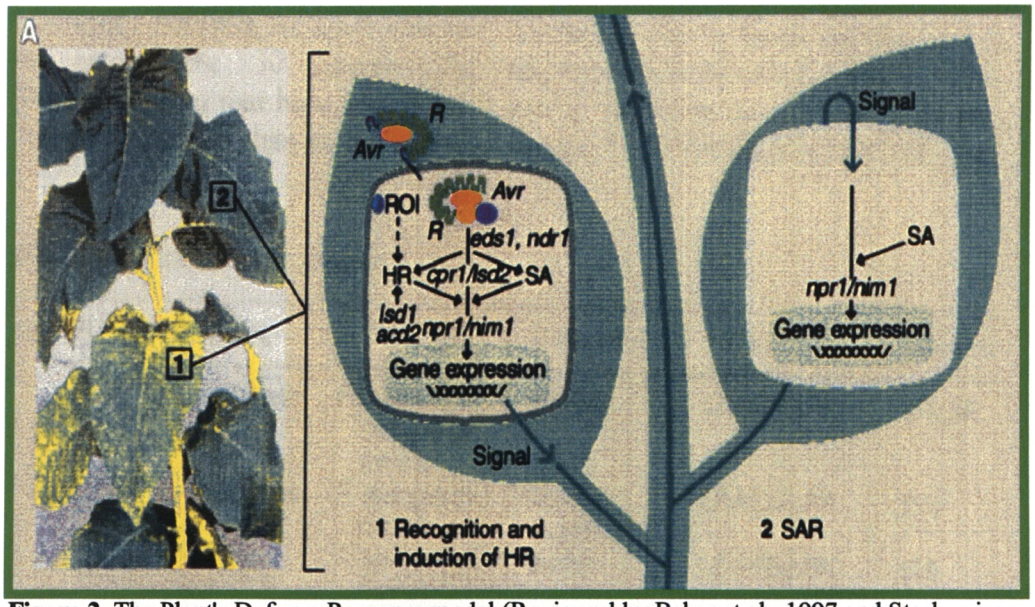


Figure 2: The Plant's Defense Response model (Reviewed by Baker et al. 1997 and Staskawicz et al. (1995) The avirulence signal is recognized by the resistance gene product either intracellularly or extracellularly(1). This interaction elicits a series of signal transduction events and to the induction of defense responses. Among these is the hypersensitive response (HR)at the site of infection (1) and the systemic acquired resistance(SAR) which is elicited throughout the plant (2) by a signal compound thought to be salycilic acid (SA).

The pathogen's avirulence gene produces an avirulence signal that comes in contact with the plant upon pathogen attack. The plant is then able to recognize the avirulence signal through direct or indirect interaction with the product of the plant's R gene. The interaction of the pathogen's avr gene and the plant's R gene then elicits a series of defense responses through signal transduction cascades. A common defense response by

the plant is the induction of a hypersensitive response or HR. The HR is a form of programmed cell death at the site of infection by the pathogen, that prevents the pathogen from spreading disease throughout the plant. Other defense responses are involved in plant defense: induction of pathogenesis related proteins (PR proteins), induction of systemic acquired resistance (SAR), production of reactive oxygen intermediates (ROI) and phytoalexins (Figure 2).

DISEASE RESISTANCE GENES

The first cloned disease resistance gene or R gene was found in 1992 by Johal and Briggs. This gene, found in maize, called the Hm1 gene, confers resistance against a fungal pathogen *Cochliobolus carbonum*. However, this gene does not follow the gene-for-gene model exactly because in the absence of a functional gene producing the toxin, the fungus is unable to colonize the plant and consequently when the pathogen is avr- the plant is "resistant".

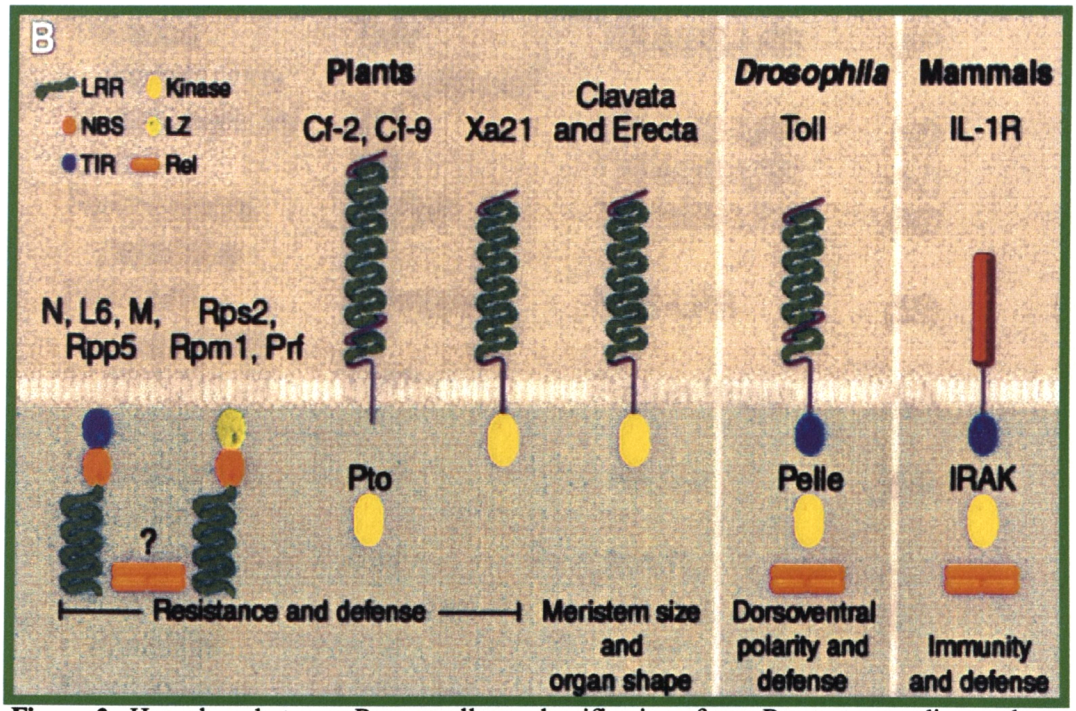


Figure 3: Homology between R genes allows classification of new R genes according to the domains in their protein structures. Homologous motifs also have been found in other plant, insect and mammalian proteins with different functions but which are probably involved in signal transduction pathways. The white band in the middle represents the cell's membranes.

To date scientists have found and cloned a great number of disease resistance genes. The first gene that followed the gene-for-gene model as described by Flor was the tomato Pto gene for resistance against the bacterial pathogen *Pseudomonas syringae* (Martin et al. 1993). A surprising finding in 1995 by Staskawicz et al. showed that significant

homologies existed between several of the R genes that had been cloned up to date. They found that the tobacco N gene, the flax L6 and the Arabidopsis RPS2 gene shared structural similarities, and showed high homology between several regions and that Cf-9, RPS2, L6 and N genes all contained leucine-rich repeat (LRR) regions. Furthermore, they found that L6, RPS2 and N all shared homology at a region known as the nucleotide binding site (NBS).

To date over twenty R genes have been cloned that follow Flor's model for resistance. These genes still are found to share homology in the regions described and are now classified according to their structure similarities (Figure 3 and Table 2).

Table 2: Classification of the most studied disease resistance genes.

Gene	R gene class	Organism	Common Name	Pathogen	Reference
N	TIR-NBS-LRR	*Nicotiana tobaccum*	Tobacco	Tobacco Mosaic Virus	Whitam et al. 1994
L6	TIR-NBS-LRR	*Linum usitatissimum*	Flax	*Melampsora lini*	Lawrence et al. 1995
M	TIR-NBS-LRR	*Linum usitatissimum*	Flax	*Melampsora lini*	Ellis et al. 1995
RPP5	TIR-NBS-LRR	*Arabidopsis thaliana*	Arabidopsis	*Peronospora parasistica*	Parker et al. 1997
RPP1	TIR-NBS-LRR	*Arabidopsis thaliana*	Arabidopsis	*Peronospora parasistica*	Bottella et al. 1998
RPM1	LZ-NBS-LRR	*Arabidopsis thaliana*	Arabidopsis	*Pseudomonas syringae*	Grant et al. 1996
RPS2	LZ-NBS-LRR	*Arabidopsis thaliana*	Arabidopsis	*Pseudomonas syringae*	Bent et al. 1994
RPP8	LZ-NBS-LRR	*Arabidopsis thaliana*	Arabidopsis	*Peronospora parasitica*	McDowell et al. 1998
Prf	LZ-NBS-LRR	*Lycospersicon esculentum*	Tomato	*Pseudomonas syringae*	Salmeron et al. 1998
Mi	LZ-NBS-LRR	*Lycospersicon esculentum*	Tomato	*Meloidogyne incognita*	Milligan et al. 1998
				Macrosiphum euphorbiae	Rossi et al. 1998
Co-2*	NBS-LRR	*Phaseolus vulgaris*	Common Bean	*Colletotrichum lindemuthianum*	Creusot et al. 1999
Pib	NBS-NBS-LRR	*Oryza sativa*	Rice	*Magnoporthe grisea*	Wang et al. 1999
I2	NBS-LRR	*Lycospersicon esculentum*	Tomato	*Fusarium oxysporum*	Ori et al. 1997
Rp1	NBS-LRR	*Zea Mays*	Maize	*Puccinia sorghi*	Collins et al. 1999
Xa1	NBS-LRR	*Oryza sativa*	Rice	*Xanthomonas oryzae*	Yoshimura et al. 1998
Cf4	LRR-TM	*Lycospersicon esculentum*	Tomato	*Cladosporium fulvum*	Jones et al. 1994
Cf9	LRR-TM	*Lycospersicon esculentum*	Tomato	*Cladosporium fulvum*	
Cf2	LRR-TM	*Lycospersicon esculentum*	Tomato	*Cladosporium fulvum*	
Cf5	LRR-TM	*Lycospersicon esculentum*	Tomato	*Cladosporium fulvum*	Dixon et al. 1996
HS1pro-1	LRR-TM	*Beta vulgaris*	Sugar beet	*Heterodera schachtii*	Cai et al. 1997
Pto	Protein Kinase	*Lycospersicon esculentum*	Tomato	*Pseudomonas syringae*	Martin et al. 1993
Xa21	LRR-Kinase	*Oryza sativa*	Rice	*Xanthomonas oryzae*	Song et al. 1995
Mlo	533aa TM	*Hordeum vulgare*	Barley	*Ersyphe graminis*	Buschges et al. 1997
Hm1	Toxin reductase	*Zea mays*	Maize	*Cocchliobolus carbonum*	Johal and Briggs 1992

* The genes belonging to the Co-2 are still candidate disease resistance genes tightly linked to anthracnose resistance. This gene was included in this table and in the alignments since it is the closest candidate of an R gene in a legume species.

DOMAINS OF RESISTANCE GENES ENCODING A NUCLEOTIDE-BINDING SITE

The table of the cloned resistance genes so far shows the classification of resistance genes in groups based on the structure similarities. The most abundant classes of R genes found so far are the classes containing a nucleotide-binding site and a leucine-rich repeat: TIR-NBS-LRR, LZ-NBS-LRR and NBS-LRR. It has been estimated from Arabidopsis studies that over 1% of a plant's genome is composed NBS-containing R genes (Meyers et al. 1999). The genes containing a nucleotide-binding site can also be classified according to the other domains present:

TIR, or Toll -Interleukin Receptor: Located in the amino terminal of several R genes, the TIR is thought be a determinant of the signal transduction pathway that is elicited after interaction of the R gene product with the avirulence signal (Aarts et al. 1995). This motif derives its name from the homology shared with the Toll receptor in Drosophila and with the Interleukin-1 Receptor in mammals. The Toll and Interleukin-1 receptors are both involved in eliciting defense responses, and the homology existing among insects, mammals and plants suggests that the existence of R genes dates back to prior to the evolutionary period when plants and animals diverged (Baker et al 1997).

LZ or Leucine Zipper: The LZ motif is generally thought to be involved in homodimerization or heterodimerization of proteins. Like TIR, it is located at the amino terminal of the R gene products and is believed to be involved in determining which signal transduction pathway is elicited. This finding was found by mutant analysis of EDS1 and NDR1 genes which have been found to act in two separate defense response pathways (Aarts et al. 1995). Surprisingly, they found that a group of the Arabidopsis R

genes tested, were dependent on the EDS1 gene for eliciting the defense response. On the contrary, a second group of Arabidopsis R genes was found to be dependent on NDR1 for eliciting the defense response. The group of genes dependent on EDS1 was found to have the TIR motif, while the genes dependent on NDR1 was found to have a LZ motif.

LRR or Leucine Rich Repeat: The LRR is a region found at the carboxyl terminal of NBS-LRR genes and at the amino terminal of other R genes. The LRR consists of several repeats of regions with leucine amino acids. This domain is found in several groups of genes: TIR-NBS-LRR, LZ-NBS-LRR, NBS-LRR, TM-LRR, Kinase-TM-LRR. The LRR region of the genes that have TM and are extracellular have more perfect repeats of leucine than the LRR of the genes that are found in the cytoplasm and which have an NBS or nucleotide-binding site. Leucine-rich repeat regions are involved in protein-protein interactions. For this reason, scientists believe that LRR's are involved in avr signal binding, and thus determine pathogen specificity. The Cf-4 and Cf-9 genes have different race specificities but differ in only a few amino acids in the LRR domain. Variation in LRR's is directly correlated in some cases with new recognition specificities. A truncated form of the Xa21 gene that lacked the kinase domain but still had the functional LRR domain maintained its original specificity (Reviewed by Martin et al, 1999).

Domain swap experiments on alleles of the L gene showed that the specificity of the recombinant gene was correlated with the specificity of the gene from which the LRR domain was obtained. Furthermore, the M gene in flax contains two single repeats in the C-terminal of the LRR region, but mutants, possessing one single repeat lost M resistance

but gained no detectable specificity (Anderson, et al 1997). There still remains more evidence to be found that confirms that the LRR's are responsible for conferring pathogen specificity.

NBS or Nucleotide-Binding Site: A major portion of the genes shown in Table 2 belong to the TIR-NBS-LRR, LZ-NBS-LRR, and the NBS-LRR groups. The genes that fall into these three groups share significant homology at the nucleotide-binding site or NBS. The NBS regions have been found in other signaling proteins including the RAS group, Adenosine Triphosphatases ATPases, elongation factors, and heterotrimeric G-proteins (Bourne et al. 1991). These proteins are critical for numerous fundamental eucaryotic cellular events such as cell growth, differentiation, cytoskeletal organization, vesicle transport and defense (Baker, 1997, and Bourne, 1991). All the genes that contain a cytoplasmic LRR have an NBS domain, which indicates that binding of GTP or ATP is likely to be essential for their function.

Since Staskawicz et al.'s findings concerning the homology of R genes, scientists have directed much of their research efforts towards looking for candidate resistance genes taking advantage of the existing homology between R genes. The Nucleotide Binding Site has been found to be the region that shares the highest homology between the well-characterized resistance genes (Figure 4).

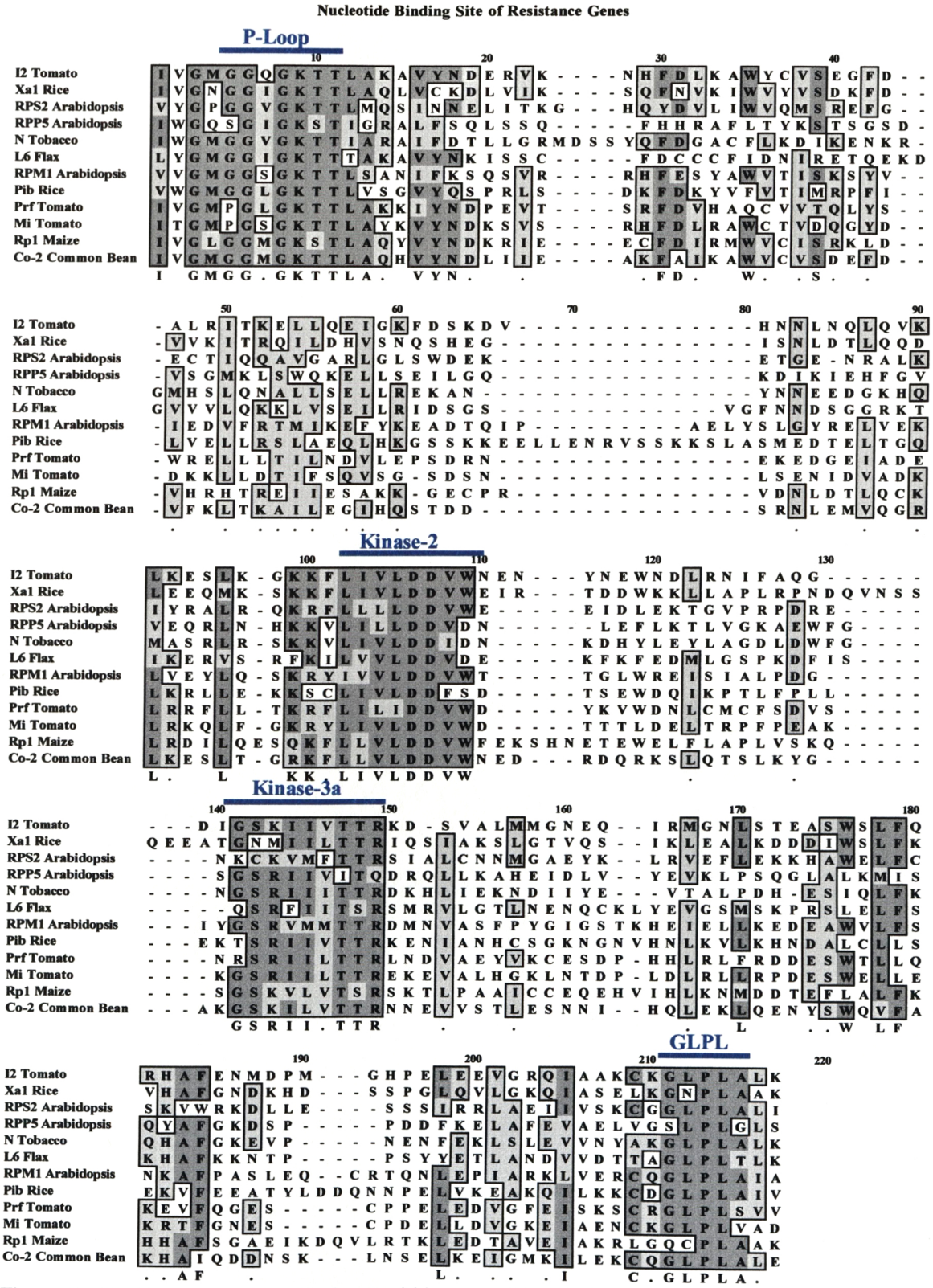


Figure 4: Highly conserved motifs found within nucleotide-binding site (NBS) region of characterized disease resistance genes: P-Loop, Kinase-2, Kinase-3a, and GLPL.

Additionally, the NBS region contains several highly conserved motifs within the already highly homologous sequence. Of those motifs, the P-Loop and the Kinase-2 domains have been well characterized for ATP- and GTP- binding proteins (Bourne et al. 1991, Traut, et al,1994, Walker et al. 1982). These motifs are very highly conserved since they have been found to play crucial roles in the functioning of the nucleotide-binding sites. The P-Loop interacts directly with the phosphate of the bound nucleotide (NTP), (Meyers et al. 1999, Saraste et al. 1990) and mutations on key residues in the P-loop of the tobacco N gene have led to partial loss of function (Baker et al. 1997). The kinase-2 domain is involved in coordinating the ion (Mg2+) required for phospho-transfer reactions (Traut et al. 1994). Additional motifs like the Kinase-3a and the GLPL or putative membrane-spanning domain have also been found in the nucleotide-binding site regions of plant resistance genes.

A comprehensive study looked at most if not all, of the NBS-containing sequences in plants which were available in the public databases at the time (Meyers et al. 1999). Their study included the NBS regions from 14 well-characterized R-genes, and 481 additional NBS sequences derived from BAC, Genomic, PCR or EST sequences from 26 plant genera. Their findings showed that the plant regions grouped in two major trees, one containing TIR-NBS genes and another containing nonTIR-NBS. In addition, they found motifs that are characteristic for all plant NBS regions and others that are conserved exclusively in TIR-NBS and others in NonTIR-NBS (Table 3).

Table 3: Characteristic motifs of the Nucleotide Binding Site in Plants.

Motif	Other names	Consensus	Location[a]
P-Loop	Kinase-1a, G-1	GVGKTT	1
	RNBS-A-nonTIR	FDLxAWVCVSQxF	20
	RNBSA-TIR	FLENIRExSKKHGLEHLQKKLLSKLL	23
Kinase-2	G-3	LLVLDDVW	74
Kinase-3a	RNBS-B	GSRIIITTRD	102
	RNBS-C	YEVxxLSEDEAWELFCKxAF	122
GLPL		CGGLPLA	162
	RNBS-D-NonTIR	CFLYCALFPED	223
	RNBS-D-TIR	FLHIACFF	219

[a]Approximate locations based on average of 14 characterized R genes from Meyers et al. 1999.

CANDIDATE DISEASE RESISTANCE GENES

The characteristics of the NBS regions, such as the correlation with disease resistance, the high abundance in plant genomes, the homology between sequences and the highly conserved motifs, allow scientists to use the NBS regions as a tool for the identification of candidate disease resistance genes.

Several methods have provided NBS sequence data to the public databases. However, the most practical and one which has proven highly reliable is the polymerase chain reaction (PCR) with a degenerate primers approach. This method takes advantage of the highly conserved motifs within the NBS regions. By creating degenerate primers for the NBS motifs, several groups have been able to amplify a representative sample of the NBS superfamily. This approach has been used successfully in soybean (Kanazin et al. 1996; Yu et al. 1996), potato (Leister et al. 1996), and lettuce (Shen et al. 1998). The P-Loop, the Kinase-3a and the GLPL motifs have been the most commonly used motifs in the degenerate PCR approach. In soybean, PCR using a forward primer designed for the P-Loop with a reverse primer for the GLPL motif (Kanazin et al. 1996); and with a reverse primer for the Kinase-3a motif (Yu et al. 1996) detected at least eleven different classes of NBS which are located in at least 11 of the 26 linkage groups in the soybean genetic map. Of the eleven classes at least five are mapped near resistance gene loci. (Kanazin et al. 1996, Yu et al. 1996). Similar results were obtained in maize where eleven classes of NBS were found of which six mapped in close proximity to disease resistance loci (Collins et al. 1998).

Based on the results obtained from the different research groups using the degenerate PCR approach we pursued a similar approach to study the existence and the variability in

the Nucleotide Binding Site family of resistance gene analogs in *Medicago sativa.* The characterization of the nucleotide-binding site super-family of disease resistance genes provided by this study will be fundamental in developing strategies for the development of disease and pest resistant alfalfa in the future.

MATERIALS AND METHODS

ISOLATION OF *MEDICAGO SATIVA* RESISTANCE GENE ANALOGS

Nucleic Acids. Genomic DNA was extracted from an alfalfa plant of cultivar UC123 using the method described by Dellaporta et al. 1983. DNA from UC123 was used for isolation of the *Medicago sativa* resistance gene analogs.

Degenerate Polymerase Chain Reaction. PCR was run on UC123 genomic DNA using primers designed from the P-Loop and the GLPLA motifs of the NBS region of R genes as described by Kanazin et al. (1996). The primers used were:

LM638: 5'-GGIGGIGTIGGIAAIACIAC-3'

REV7: 5'-ARIGCTARIGGIARICC-3'

Each PCR cycle consisted of 94° for 1 min, 45° for 30s and 68° for 30s. Amplification products were visualized in a 1% agarose gel with ethidium bromide. Amplification products were digested with Tsp5091 (↓AATT) (New England Biolabs, Beverly, MA) and separated using polyacrylamide gel electrophoresis (PAGE).

Cloning of the NBS regions was done using the Topo TA cloning kit (Invitrogen, Carlsbad, CA) with 100ng of pooled NBS PCR amplifications. White colonies were selected and DNA was extracted using the boiling miniprep method. Clones were confirmed by EcoRI (Promega, Madison, WI) restriction digests following instructions from the manufacturer, and visualized on a 1% agarose gel.

Individual clones were digested with Tsp509I (↓AATT) enzyme and digests were separated using PAGE gels. Ninety-four individual clones were PCR amplified using M13 primers. Amplification products from each clone were purified using Centricon

YM-100 spin columns (Millipore, Bedford, MA) prior to sequencing in order to remove unincorporated primers and nucleotides.

Southern Blots. EcoRI digests from selected white colonies were separated in a 0.8% agarose gel using Lamba DNA digested with BglI as a size marker. DNA was blotted on a MAGNA nylon transfer membrane (Micron separations, Westboro, MA). Five PCR products amplified with M13 primers that showed different restriction patterns with Tsp509I were used as probes for the NBS regions. Lambda phage DNA was used as a control probe for the L-BglI marker. Probes were labeled with $dCTP^{32}$ using the Prime-It Random Primer Labeling kit (Stratagene, LaJolla, CA). Probes were cleaned using Nuc Trap Purification columns (Stratagene, LaJolla, CA) in order to remove unincorporated radioactive nucleotides and random primers. Blots were allowed to hybridize at 42°C overnight.

Sequencing reactions were prepared using the Perkin Elmer Big Dye DNA sequencing kit (Perkin Elmer, Foster City, CA). Sequencing reactions were cleaned using the Centri-sep spin columns (Princeton Separations, Adelphia, NJ). Sequencing was done at the DNA Sequencing and Synthesis Facility at Iowa State University, Ames, Iowa.

Sequence analyses and alignments were done using MacVector 6.5 (Oxford Molecular, Madison, WI) and Assemblylign (Oxford Molecular, Madison, WI) software. Sequence comparisons with NBS from other species were done using the BLAST algorithm on the GenBank database.

Bootstrap analysis was performed using the neighbor-joining method and the unweighted pair-group method using arithmetic averages method using PAUP software (Sinauer Associates, Sunderland, MA).

GenBank submissions. Sequences for unique sequences obtained were submitted to the GenBank database using the Sequin software. Accession numbers for each unique alfalfa NBS sequence are the following:

Table 4: GenBank Accession numbers of resistance gene analogs (RGA) from *Medicago sativa.*

RGA	Accession number
MSRGA1	AF230813
MSRGA10	AF230814
MSRGA11	AF230815
MSRGA12	AF230816
MSRGA13	AF230817
MSRGA14	AF230818
MSRGA15	AF230819
MSRGA16	AF230820
MSRGA17	AF230821
MSRGA18	AF230822
MSRGA19	AF230823
MSRGA2	AF230824
MSRGA20	AF230825
MSRGA21	AF230826
MSRGA22	AF230827
MSRGA23	AF230828
MSRGA24	AF230829
MSRGA25	AF230830
MSRGA27	AF230832
MSRGA28	AF230833
MSRGA29	AF230834
MSRGA3	AF230835
MSRGA35	AF230836
MSRGA36	AF230837
MSRGA37	AF230838
MSRGA38	AF230839

RGA	Accession number
MSRGA39	AF230840
MSRGA4	AF230841
MSRGA40	AF230842
MSRGA41	AF230843
MSRGA42	AF230844
MSRGA43	AF230845
MSRGA44	AF230846
MSRGA46	AF230848
MSRGA47	AF230849
MSRGA48	AF230850
MSRGA49	AF230851
MSRGA5	AF230852
MSRGA50	AF230853
MSRGA51	AF230854
MSRGA52	AF230855
MSRGA53	AF230856
MSRGA54	AF230857
MSRGA55	AF230858
MSRGA56	AF230859
MSRGA6	AF230860
MSRGA7	AF230861
MSRGA8	AF230862
MSRGA9	AF230863
MSRGA57	AF230864
MSRGA58	AF230865

ISOLATION OF EXPRESSED *MEDICAGO SATIVA* RESISTANCE GENE ANALOGS

Nucleic Acids. Total RNA was extracted from a single alfalfa plant, ISC 35, using the TRIazol Reagent (GibcoBRL, Gaithesburg, MD) and following the manufacturer's recommended procedure. First strand cDNA was obtained by using the RETROscript RT-PCR kit (Ambion, Austin, TX).

Degenerate Polymerase Chain Reaction. PCR was run on ISC35 cDNA using the degenerate primers LM638 and REV7 mentioned earlier, and the following non-degenerate primers:

FOR-NBS: 5'-GGGGGGGTGGGGAAGACGAC-3'

REV-NBS7: 5'-AGGGCTAGGGGGAGGCC-3'

Each PCR cycle consisted of 94° for 1 min, 45° for 30s and 68° for 30s.

Amplification products were visualized in a 1% agarose gel with modified TAE buffer and stained with ethidium bromide. The 500 bp amplification product was removed from the gel using Ultrafree-DA columns (Millipore, Bedford, MA).

Cloning. Recovered DNA from the agarose was cloned using the pGEM-T Easy vector system (Promega, Madison, WI) and transformed into TOP10 cells (Invitrogen, Carlsbad, CA). Forty-eight clones were PCR amplified using M13 primers. Individual clones were digested with Tsp509I enzyme and digests were separated using PAGE gels. Inserts from clones amplified with M13 primers were selected for sequencing to confirm whether the cDNA clones were NBS regions. Amplification products from each clone were purified using Centricon YM-100 spin columns (Millipore, Bedford, MA) prior to sequencing in order to remove unincorporated primers and nucleotides.

Sequencing reactions were prepared using the Perkin Elmer Big Dye DNA sequencing kit (Perkin Elmer, Foster City, CA), and were cleaned using the Centri-sep spin columns (Princeton Separations, Adelphia, NJ). Sequencing of the cDNA clones was done at the USDA Genomics DNA Sequencing Lab at Kansas State University, Manhattan, KS.

Sequence analyses were done using MacVector 6.5 (Oxford Molecular, Madison, WI) and Assemblylign (Oxford Molecular, Madison, WI) software.

Nucleic Acids. DNA from Riley117A, *M. constricta, M. blancheana, M. radiata, M. praecox, M. platycarpa, M. medicaginoides, M. ruthenica, M. murex, M. rugosa, M. rigidula, M. orbicularis, M. polymorpha,* and *M. lupulina* was extracted using the method by Dellaporta et al. 1983. DNA from these plants was used to analyze the diversity in resistance gene analogs based on restriction patterns.

Degenerate Polymerase Chain Reaction. PCR was run on UC123, Riley117A, *M. constricta, M. blancheana, M. radiata, M. praecox, M. platycarpa, M. medicaginoides, M. ruthenica, M. murex, M. rugosa, M. rigidula, M. orbicularis, M. polymorpha, and M. lupulina* genomic DNA using primers designed from the P-Loop and the GLPLA motifs of the NBS region of R genes as described by Kanazin et al. 1996.

LM638: 5'-GGIGGIGTIGGIAAIACIAC-3'

REV7: 5'- ARIGCTARIGGIARICC-3'

Each PCR cycle consisted of 94° for 1 min, 45° for 30s and 68° for 30s.

Amplification product was visualized in a 1% agarose gel with ethidium bromide. Resistance gene analog variation was assessed with restriction digests Tsp5091 (New England Biolabs, Beverly, MA) of the 500 bp degenerate PCR amplification product. Digests were separated using polyacrylamide gel electrophoresis (PAGE).

RESULTS AND DISCUSSION

PCR with the degenerate primers designed for the NBS region of R genes showed what seemed to be a single 500 bp product on a 1% agarose gel (Figure 5). The molecular size marker used for comparison was lambda phage DNA digested with Bgl I restriction enzyme.

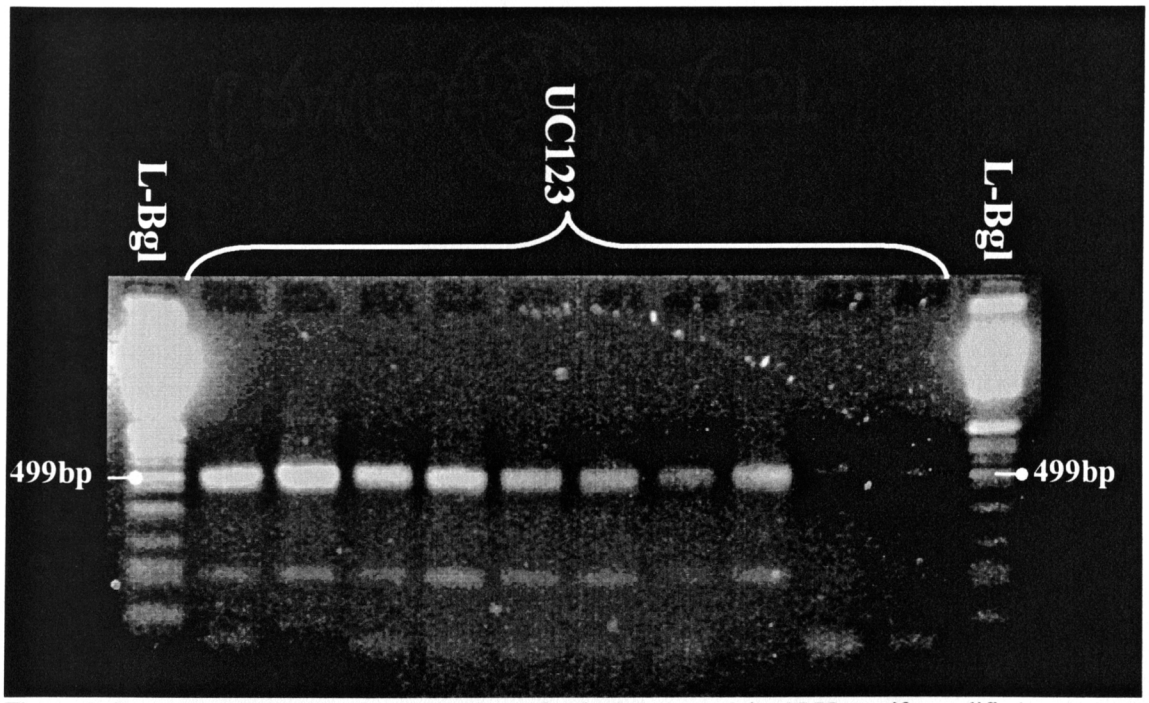


Figure 5: Degenerate PCR using primers designed for the P-Loop and the GLPL motifs amplified a 500 bp product. The molecular size standard is Lambda phage digested with BglI (L-Bgl).

However, when the original 500 bp product was digested with Tsp509I, a restriction enzyme with a four base recognition sequence, the resulting band sizes added up to more than 500 bp (Figure 6). This indicated that the original 500 bp product contained more than one single amplification product, and probably contained several different sequences that shared the same size of 500bp.

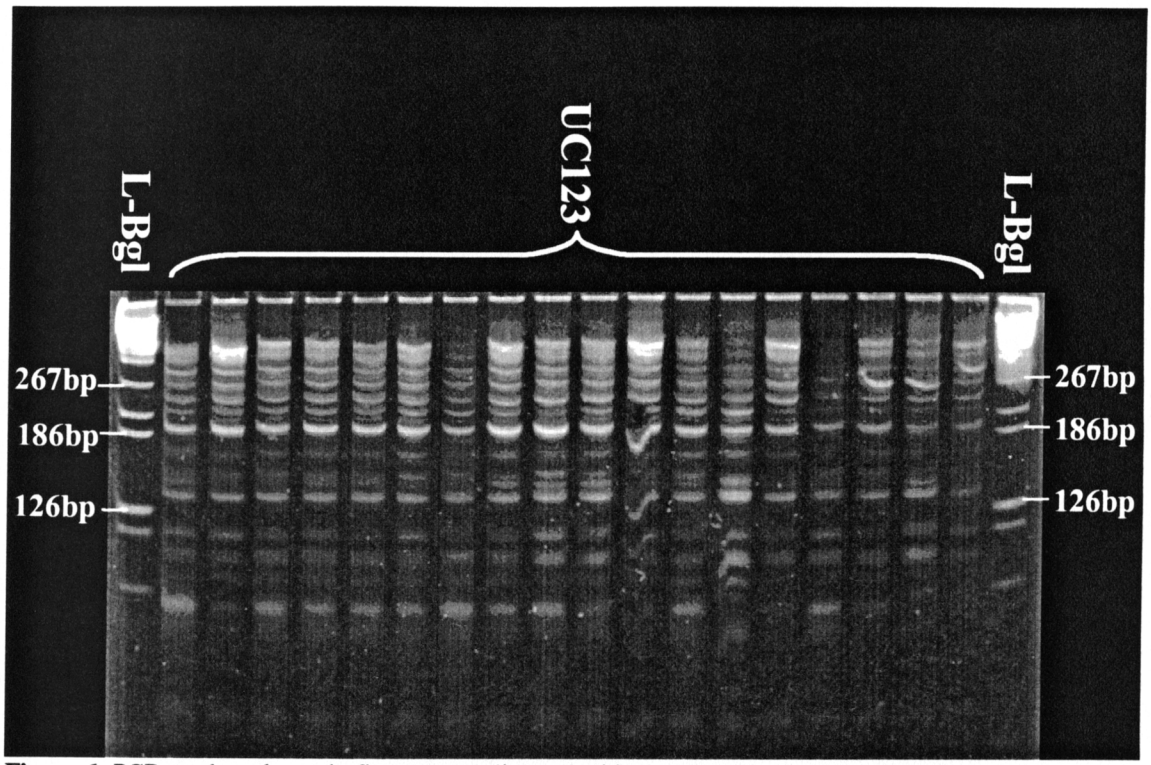


Figure 6: PCR product shown in figure 5 was digested with Tsp509I. The digest produced bands that together added up to more than the expected 500 bp for a single product. This confirmed the existence of sequence variation within the original degenerate PCR product of 500bp.

Cloning of the original 500 bp product allowed separation of the 500 bp product into individual sequences. PCR amplification of the individual clones using M13 primers gave amplification products that were in the expected ~718 bp range which takes into account plasmid sequence, ~218 bp, plus ~500 bp of NBS sequence. Although the majority of the PCR products were uniformly in the 718 bp range, a few amplification products showed smaller sized products ~ 300 bp and several had larger products, up to 1600 bp (Figure 7). 94 of the original 125 clones showed good amplification with the M13 primers.

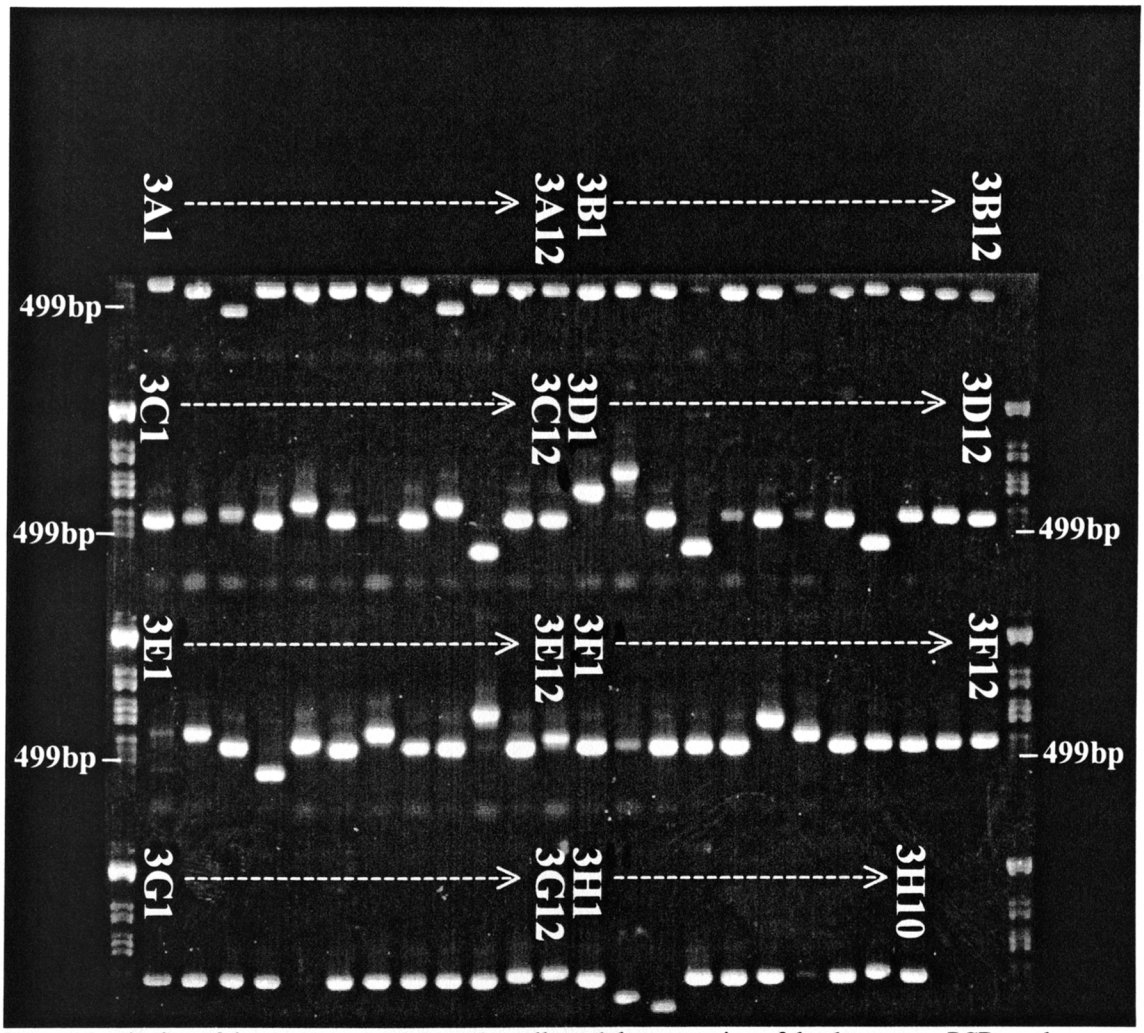


Figure 7: Cloning of the degenerate PCR product allowed the separation of the degenerate PCR product into individual sequences. Individual clones amplified with M13 primers show the expected 740 bp size. However, clones which deviated from the 740 bp size also were found. Bands that deviated from the expected 740 bp size later were found to be concatemers or shorter non-NBS sequences.

A southern blot was prepared which contained all 94 amplification products. Five individual clones were used as probes and labeled with P^{32} using the Prime-It labeling kit by Stratagene. Southern hybridization showed that all 94 clones hybridized with the selected probes indicating that there was significant homology (Data not shown). The Southern blots were non-informative in regards to separating clones that did not cross-

hybridize. For this reason, it was decided to compare the individual clones based on restriction patterns from digests with Tsp509I restriction enzyme.

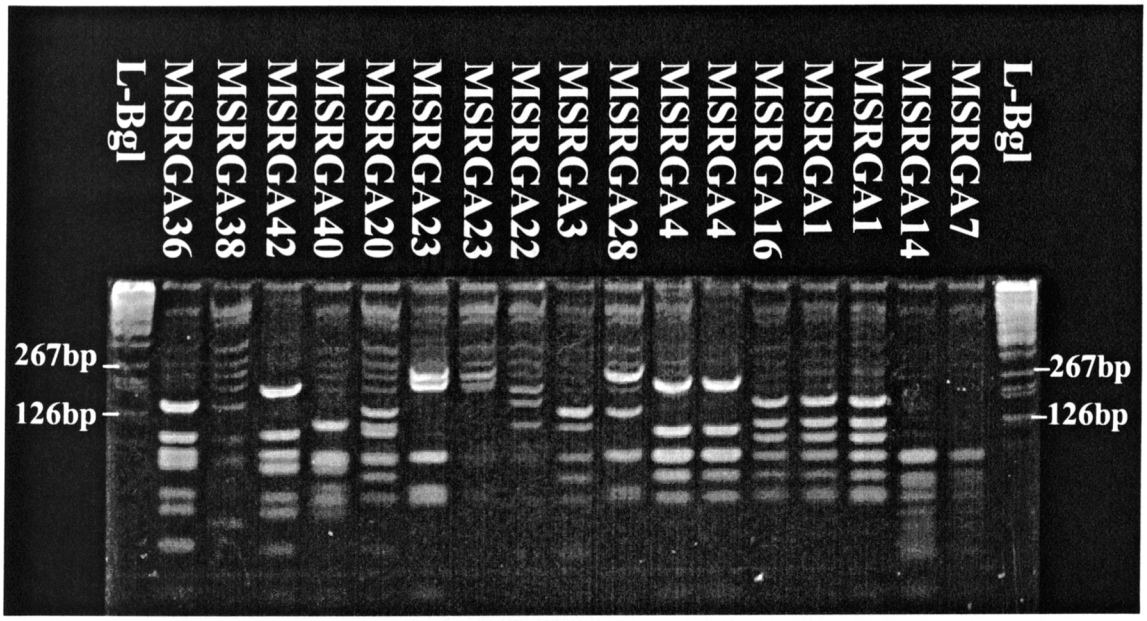


Figure 8: Individual clones digested with Tsp509I showed sequence variation between them.

A digest of the individual clones was able to show that there was indeed sequence variation within the 500 bp PCR product and between the individual clones (Figure 8). Most of the clones showed different restriction patterns and so we decided to sequence all 94 clones.

MEDICAGO SATIVA RESISTANCE GENE ANALOG NUCLEOTIDE SEQUENCES

Sequencing then allowed us to examine the NBS variation in detail (Figure 9). The 94 clones comprised 51 sequences (Table 5) which showed close homology with NBS regions from other species found in the GenBank database. The highest levels of significance obtained were found with the soybean NBS regions.

The 51 unique sequences obtained were aligned using the Clustal W algorithm through MacVector 6.5 software. The ClustalW alignments provided a dendrogram showing the genetic distance among the 51 cloned sequences (Figure 10). Bootstrap analysis using PAUP (Sinauer Associates, Sunderland, MA) provided statistical data to test the strength of the cluster assignments. The two algorithms used, unweighted pair-group method using arithmetic averages (UPGMA) and neighbor-joining (NBJ) yielded similar results, but the bootstrap analysis showed stronger support at the neighbor-joining tree (Figure 10).

The genetic distance dendrogram showed six major branches that were considered as separate clusters of MSRGA clones (Figure10). Five of the six clusters (A, B, D, E, and F) contained from 8 to13 unique MSRGA's. Although cluster C contained only one sequence, MSRGA22, it was considered as a separate cluster due to the genetic distance to the other five clusters (Figure 10). Cluster C (MSRGA22) only shares 73% homology with the closest clusters E and F (Table 6).

The six major clusters, labeled in figure 10 as A, B, C, D, E and F show between 38% to 78% identity at the nucleotide level (Table 6). However, the structure of the dendrogram shows the occurrence of subgroupings within several of the clusters. This

was observed in clusters B and E, which contain sequences that differ up to 25% and 29% (75% and 71% homology), respectively. For example, in cluster B, MSRGA48 and MSRGA36 have 75% nucleotide identity (Table 6). Likewise, in cluster E, MSRGA3 and MSRGA4 only share 71% identity. On the contrary, clusters A, D, and F show very little divergence among the several sequences found within each cluster. The sequences, within each of the clusters A, D and F share more than 95% homology. Although some of the 51 sequences showed very high homology, the few nucleotide differences might be shown to have functional importance. For example, an NBS region in potato associated with virus resistance had only 12 nucleotide differences with homologous regions found in all of the susceptible lines (Sorri et al. 1999).

For analysis purposes, families of sequences that shared more than 95% homology are represented by only one sequence of the family in order to take into account multiple alleles or PCR errors. This threshold would take into consideration possible misincorporation of nucleotides during PCR. The 5% threshold would permit 25 out of 500 misincorporated bases, which is a conservative estimate. This arbitrary parameter was used as an indicator to separate possible families of sequences. As a consequence Figure 10 shows the values for genetic distances of only those sequences which were more than 5% distant.

For example, the sequences in cluster F are represented by MSRGA7, and the sequences in cluster D are represented by MSRGA1, in Table 6 and other figures. The sequences within cluster A are represented by MSRGA23 and MSRGA28 since bootstrap analysis showed support, 75%, for two separate groups (Figure 10). In all the other clusters, sequences that differed by more than 5% were considered as different families.

This way we see that approximately 18 families of MSRGA's exist in alfalfa: MSRGA23, MSRGA28, MSRGA35, MSRGA5, MSRGA36, MSRGA40, MSRGA41, MSRGA42, MSRGA52, MSRGA48, MSRGA22, MSRGA1, MSRGA3, MSRGA55, MSRGA6, MSRGA4, MSRGA50, and MSRGA7.

The data provided from this study was based in a major part on sequence data and dendrogram analyses. In consequence, the conclusions reached regarding cluster and family assignments are highly based on homology and not on genomic map position. Map-based cloning of the *Medicago sativa* resistance gene analogs would require a fine structure recombination map and closely linked DNA markers, and genomic libraries, none of which is available for alfalfa. The autotetraploid nature of alfalfa, inbreeding depression and large genome size have hindered the development of a complete map. However, efforts on mapping the MSRGA regions characterized in this study are being carried out in collaboration with Iowa State University using an RFLP map in diploid alfalfa (Brummer et al. 1993). Additionally, research interest has increased on the alfalfa relative *Medicago truncatula* as the model research plant for legume species. This will contribute additional tools useful in alfalfa research, including an already developed BAC library of *Medicago truncatula* (Nam et al. 1999) which might facilitate the cloning of R genes useful in alfalfa transformation for disease and pest resistance.

The data obtained from this study will be useful in the isolation of complete resistance genes from alfalfa as well as other closely related species. Also, the use of other molecular techniques for the identification of markers for disease or pest resistance associated with the nucleotide binding site will likely increase the efficiency of marker assisted selection.

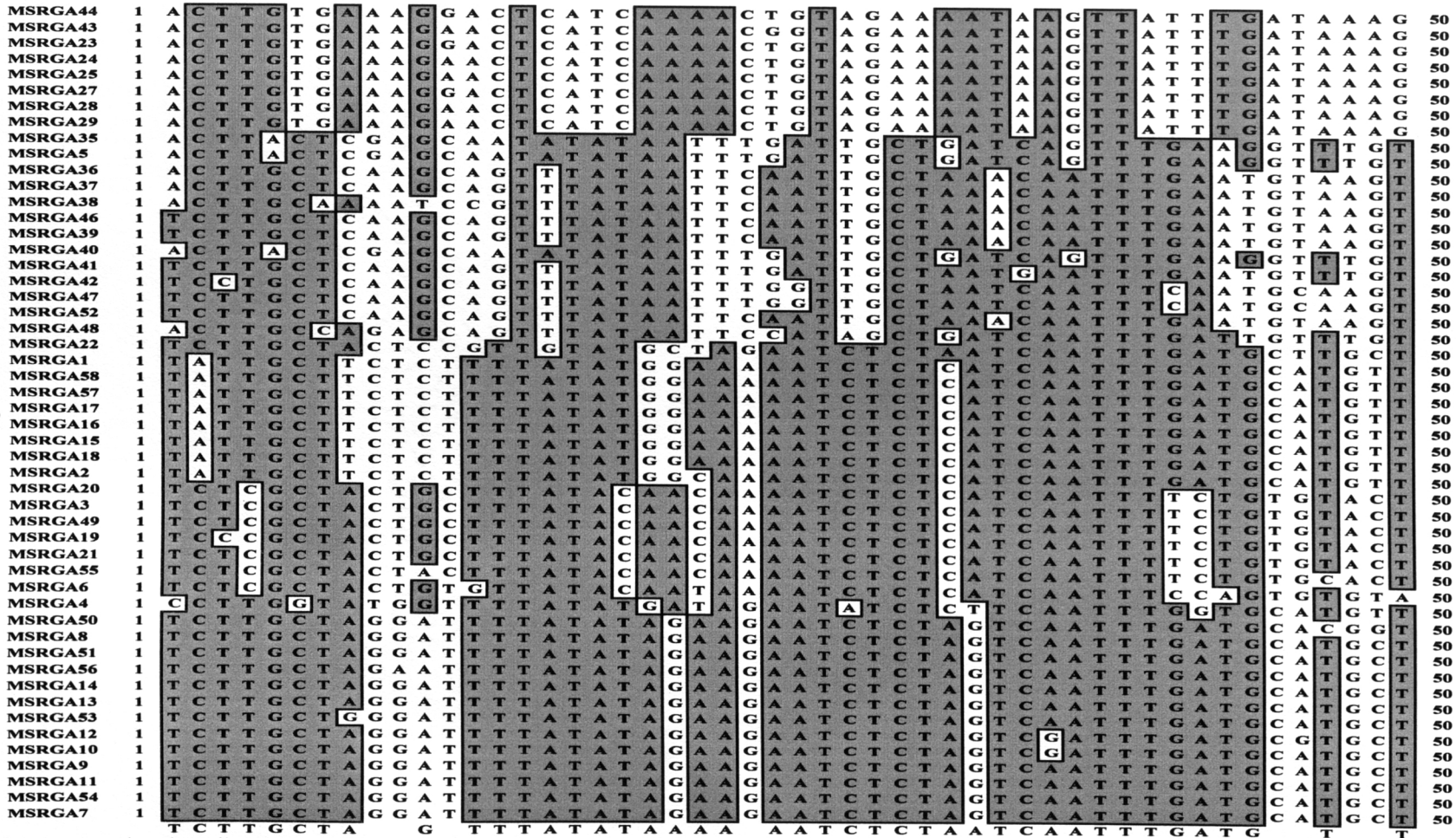
Table 5: Classification of the ninety-four clones sequenced.

Family	Size	Clones	Family	Mutation
MSRGA1	475	3A11,3B5,3C2*,3C3*,3D3,3D5,3D10*, 3D11*,3F7*,3F8,3G5*,3H1,3H6,3H8	1	
MSRGA2**	453	3E6	1	22bp DELETION & 22.236.237 244.245.247.254.264.267
MSRGA15	475	3F5	1	153.262.420
MSRGA16	475	3G7	1	431
MSRGA17	475	3F10	1	166
MSRGA18	475	3B7*	1	162.397
MSRGA57	475	3F1	1	181.219
MSRGA58	475	3G9	1	124.341
MSRGA3	472	3C9*,3H4	3	
MSRGA19	472	3A10*	3	3.142.217.445.
MSRGA20	472	3C4	3	74.217.224.312.445
MSRGA21	472	3H9*	3	14.217.301.303.445
MSRGA49	472	3D7*	3	108.217.444
MSRGA4	475	3B2,3B10	4	
MSRGA5**	469	3A4,3B6,3B8,3B9,3D6,3E2*,3E11	5	
MSRGA6	472	3B3	6	
MSRGA7	472	3G8	7	
MSRGA8	472	3F3	7	273.281.307
MSRGA9	472	3G3	7	428.32
MSRGA10	472	3C11	7	36.91.169
MSRGA11	472	3F11*	7	72
MSRGA12	472	3E8	7	36.46
MSRGA13	472	3G10	7	91.267.281
MSRGA14	472	3B12,3G11*	7	91.267.281.307
MSRGA51	472	3B11	7	270.284.310.451
MSRGA53	472	3F4	7	8.267.281.384
MSRGA54	472	3E9	7	72.135.221.307.421
MSRGA56	472	3F2*	7	10.267.281.307.420
MSRGA22	472	3E5*	22	
MSRGA23	469	3C12,3D8,3D12,3H7,3E7*	23	
MSRGA24	469	3G4	23	12
MSRGA25	469	3E3	23	12.23
MSRGA27	469	3A1*	23	73.289
MSRGA28	469	3C1,3C6	23	12.103.134.140.206.250.252 259.293.308
MSRGA29	469	3G12*	23	12.250.252.259.293.308
MSRGA43	469	3C8	23	25.103.200.268
MSRGA44**	468	3A7	23	103.134.140.176.206.366
MSRGA35**	465	3A12	35	
MSRGA36	469	3G6	36	
MSRGA37	469	3A8*	36	52
MSRGA38	469	3B1	36	7.8.11.13.356.397.410.470
MSRGA39	469	3F9*	36	1.130.144.356.368.373.397.410
	469			415.430.439.444.455.468.469
MSRGA41	469	3G2	41	
MSRGA46	469	3A6	36	122.144.254.354.356.410
MSRGA40	469	3G1	40	
MSRGA42**	465	3A2	42	
MSRGA47**	465	3C5*	42	3.95
MSRGA52**	465	3H5	52	
MSRGA48	469	3B4*	48	
MSRGA50	472	3A5	50	
MSRGA55**	474	3E12*	55	
MSAGC32	159	3H2*,3D9*,3A9*		Short sequences
MSSRD59	109	3D4*		Short repetitive DNA homology
Other***		3A3,3D1,3D2,3C7,3F12,3C10, 3E4,3E1,3E10,3F6,3H3,3H10		

* Indicates chains of primer dimers at either end of the NBS sequences affecting banding patterns with Tsp509I.

** Indicates clones with stop codons in the sequence.

*** Indicates clones with irrelevant data: No amplification, not good sequence data, or short sequences.


MSRGA44 1 ACTTGTGGAAAAAAGGGGAAACCCCTTT

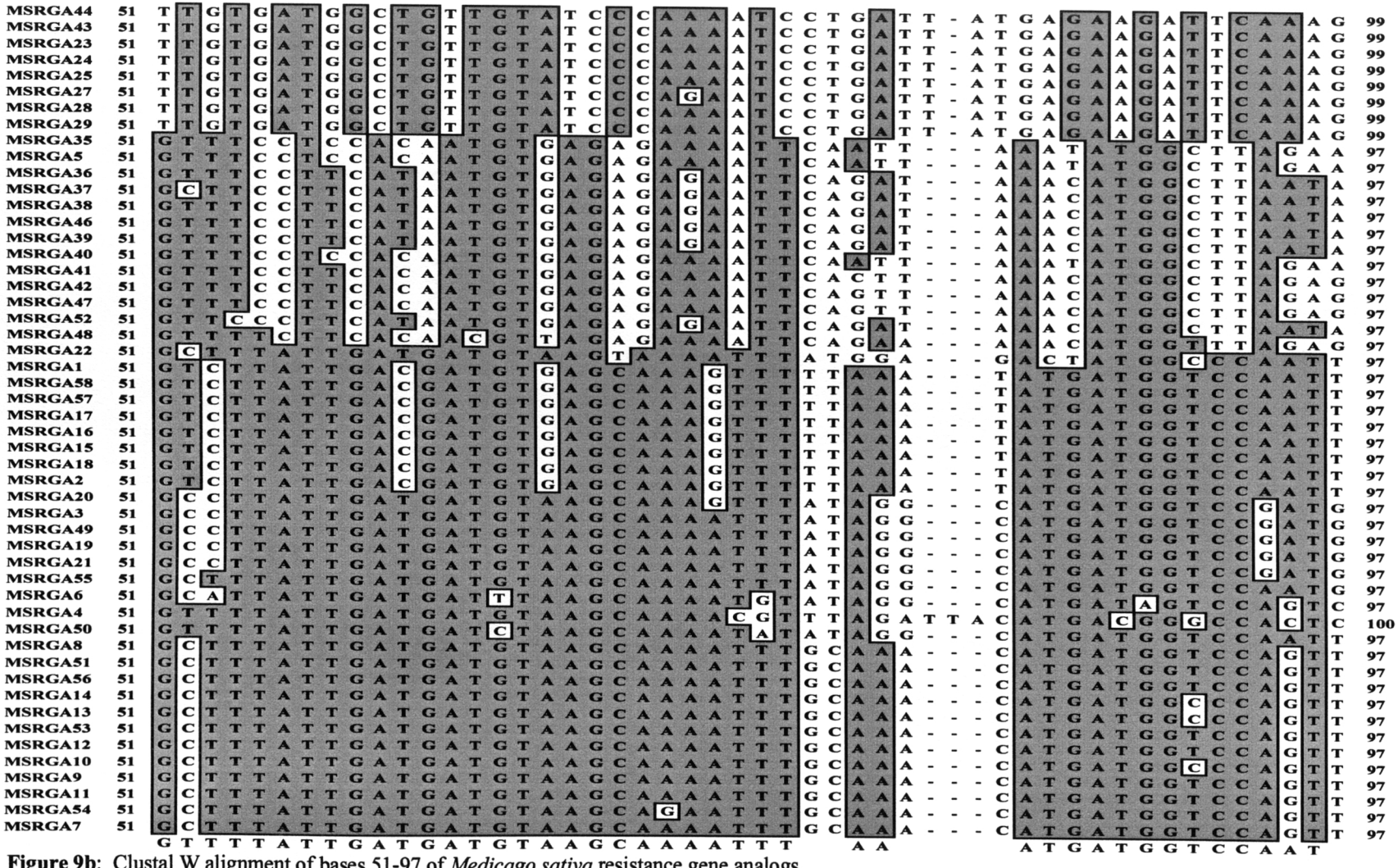
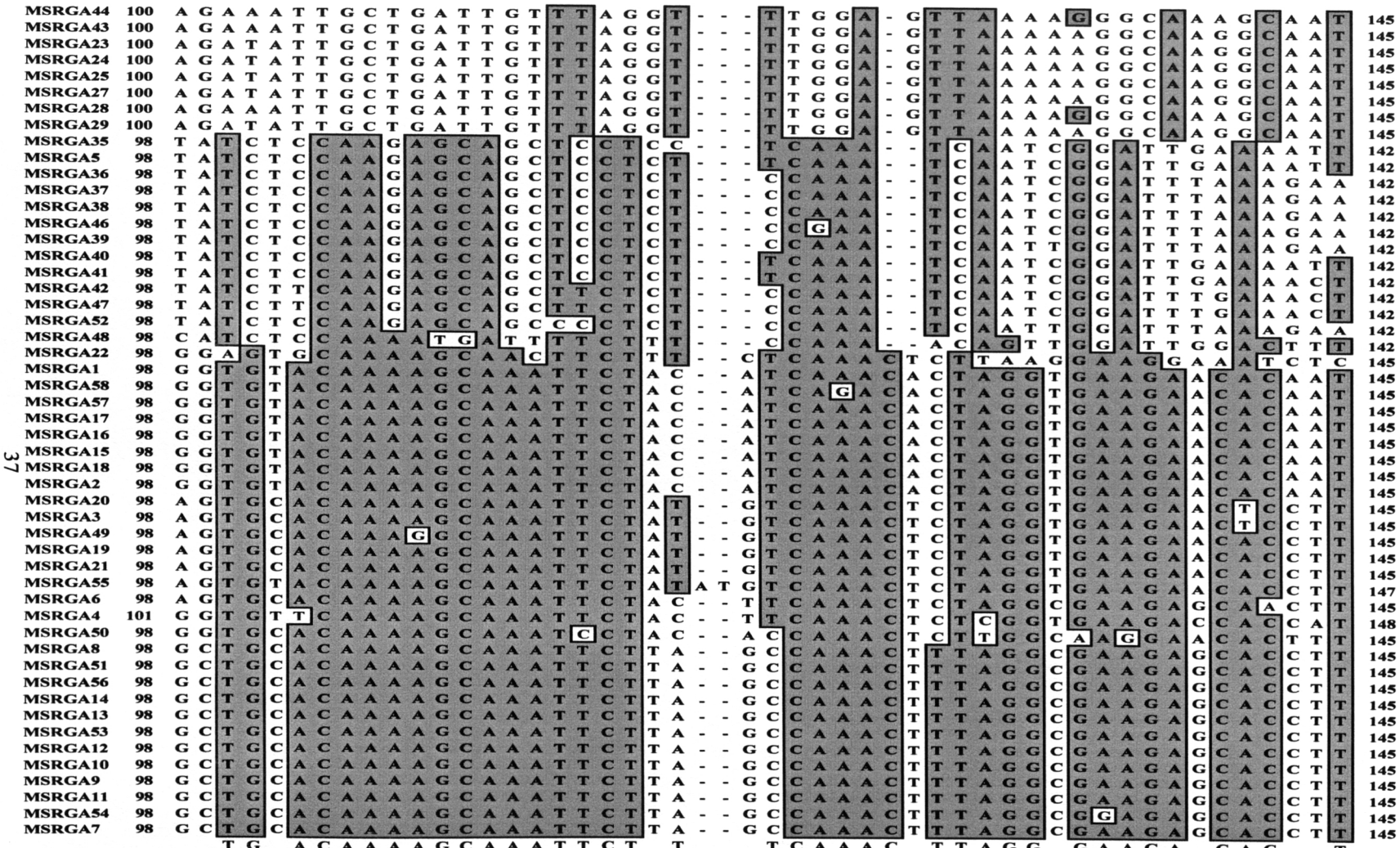


Figure 9b: Clustal W alignment of bases 51-97 of *Medicago sativa* resistance gene analogs


MSRGA44 100
MSRGA43 100
MSRGA23 100
MSRGA24 100
MSRGA25 100
MSRGA27 100
MSRGA28 100
MSRGA29 100
MSRGA35 98
MSRGA5 98
MSRGA36 98
MSRGA37 98
MSRGA38 98
MSRGA46 98
MSRGA39 98
MSRGA40 98
MSRGA41 98
MSRGA42 98
MSRGA47 98
MSRGA52 98
MSRGA48 98
MSRGA22 98
MSRGA1 98
MSRGA58 98
MSRGA57 98
MSRGA17 98
MSRGA16 98
MSRGA15 98
MSRGA18 98
MSRGA2 98
MSRGA20 98
MSRGA3 98
MSRGA49 98
MSRGA19 98
MSRGA21 98
MSRGA55 98
MSRGA6 98
MSRGA4 101
MSRGA50 98
MSRGA8 98
MSRGA51 98
MSRGA56 98
MSRGA14 98
MSRGA13 98
MSRGA53 98
MSRGA12 98
MSRGA10 98
MSRGA9 98
MSRGA11 98
MSRGA54 98
MSRGA7 98
145
145
145
145
145
145
145
145
142
142
142
142
142
142
142
142
142
142
142
142
142
145
145
145
145
145
145
145
145
145
145
145
145
145
145
147
145
148
145
145
145
145
145
145
145
145
145
145
145
145
145
145

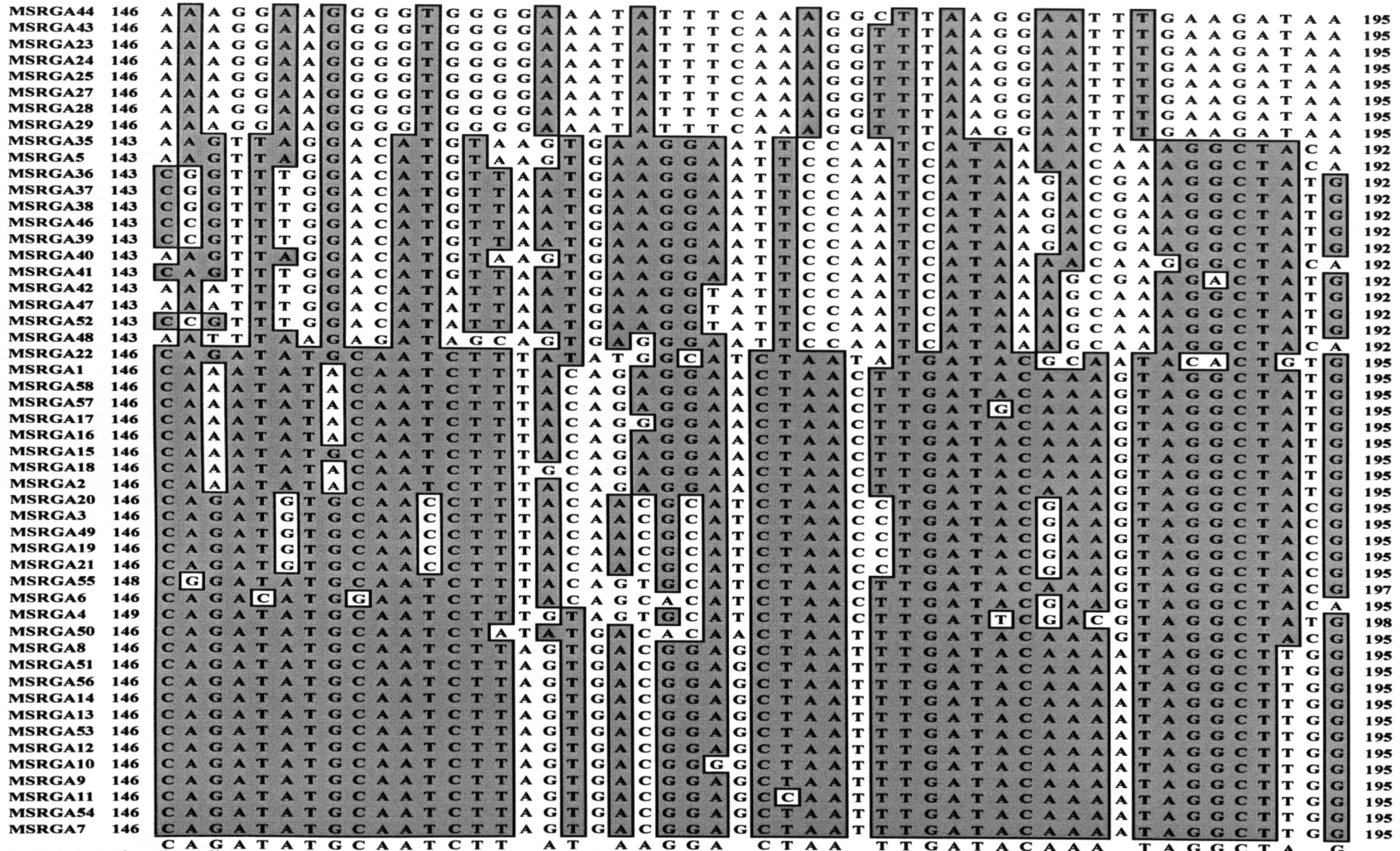
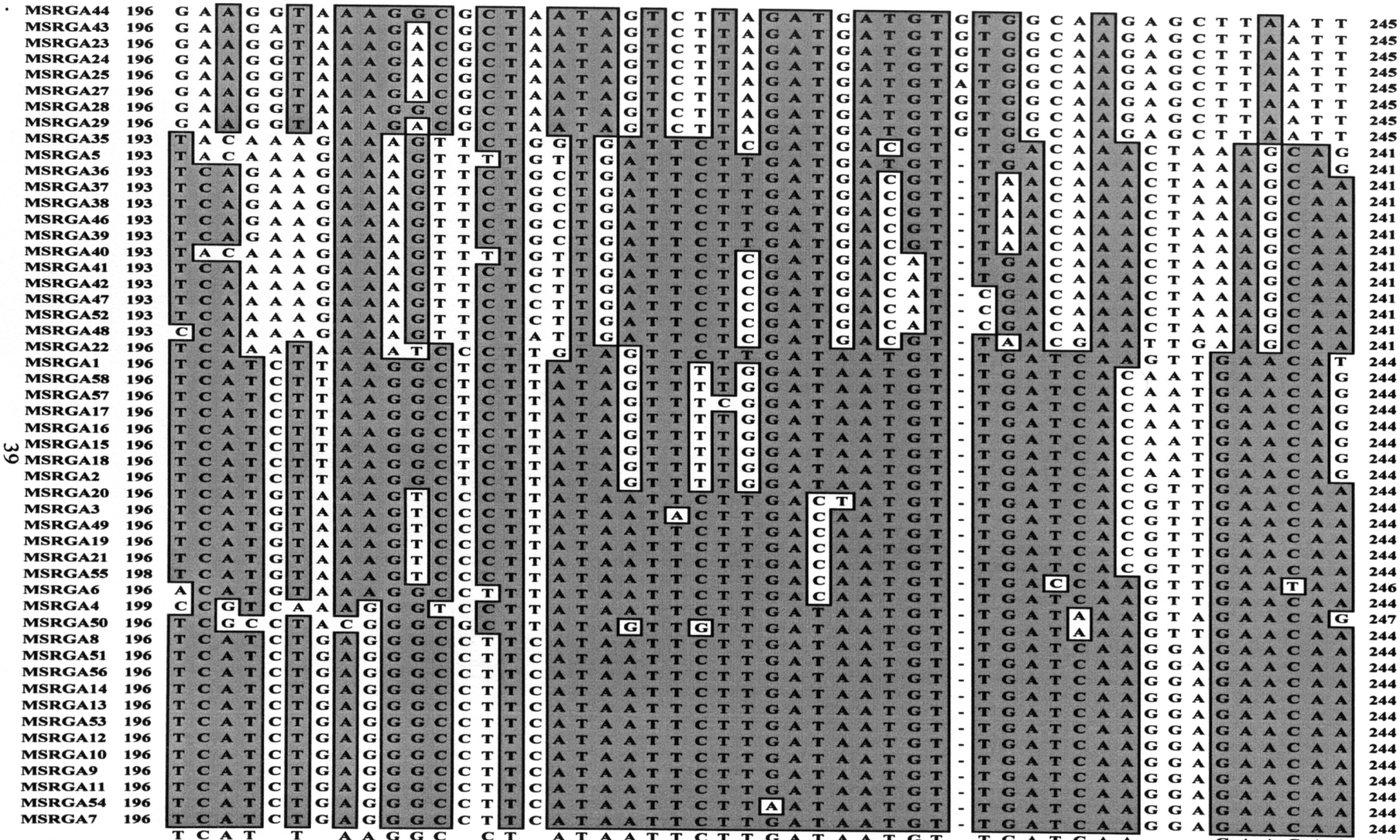
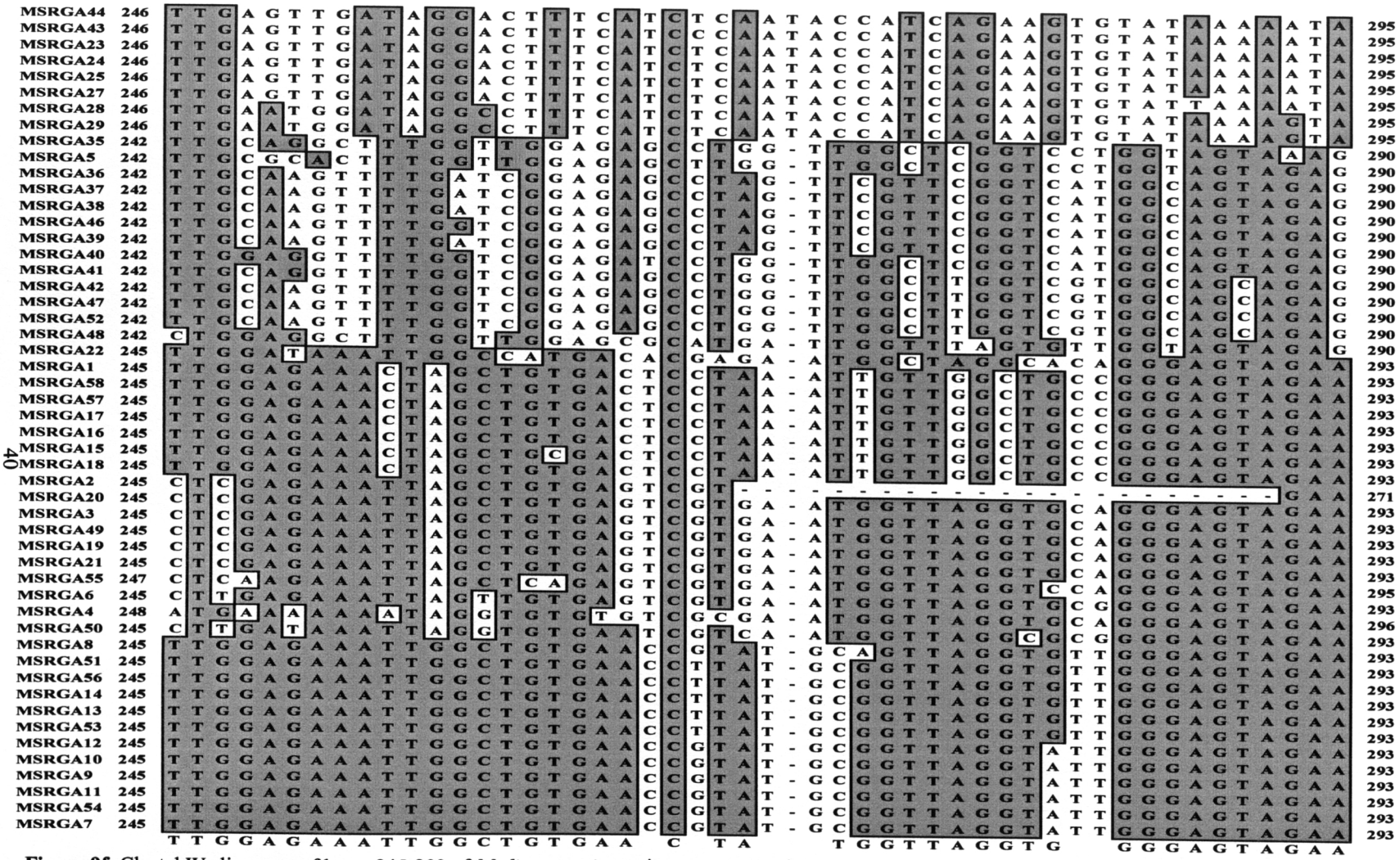
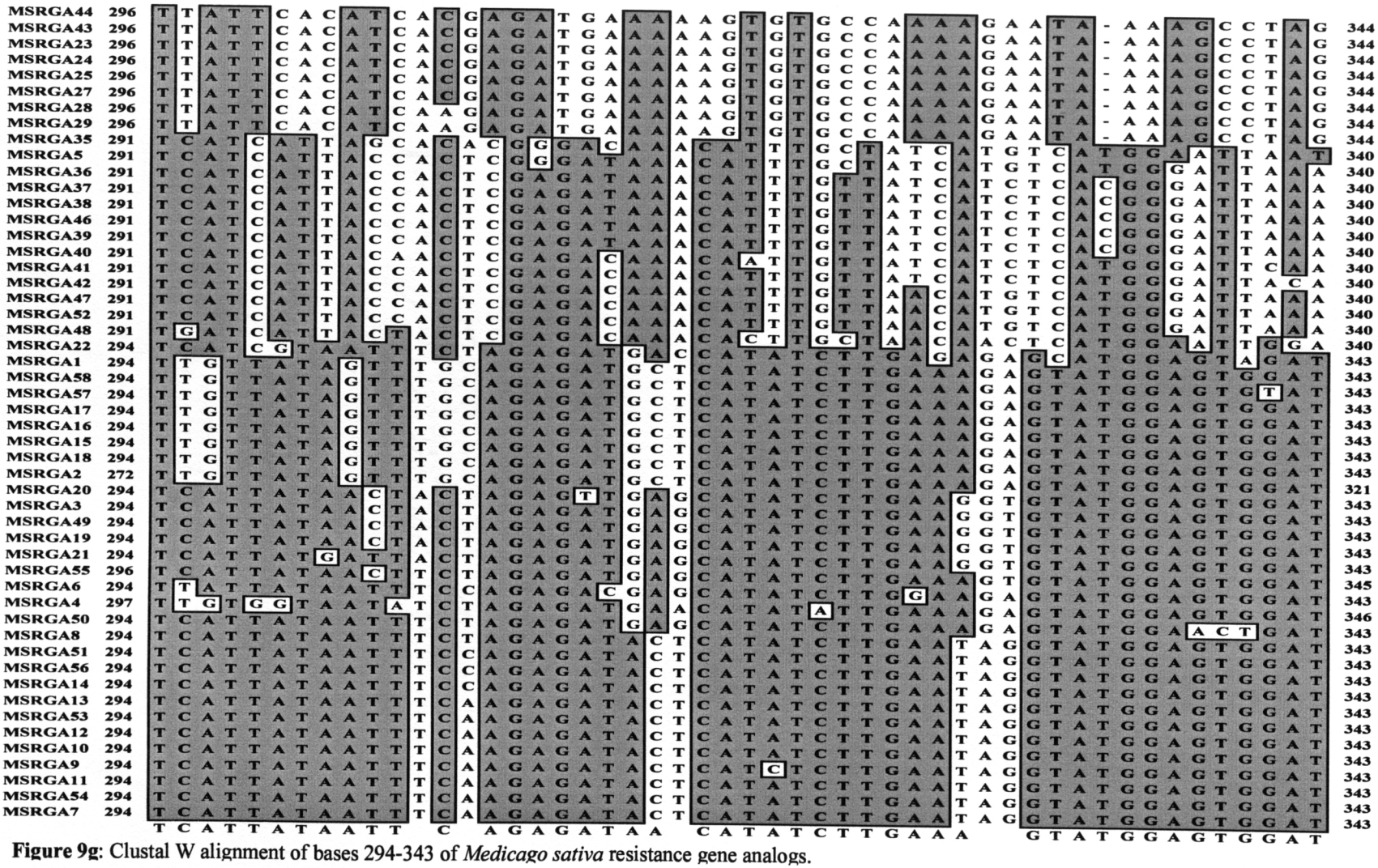
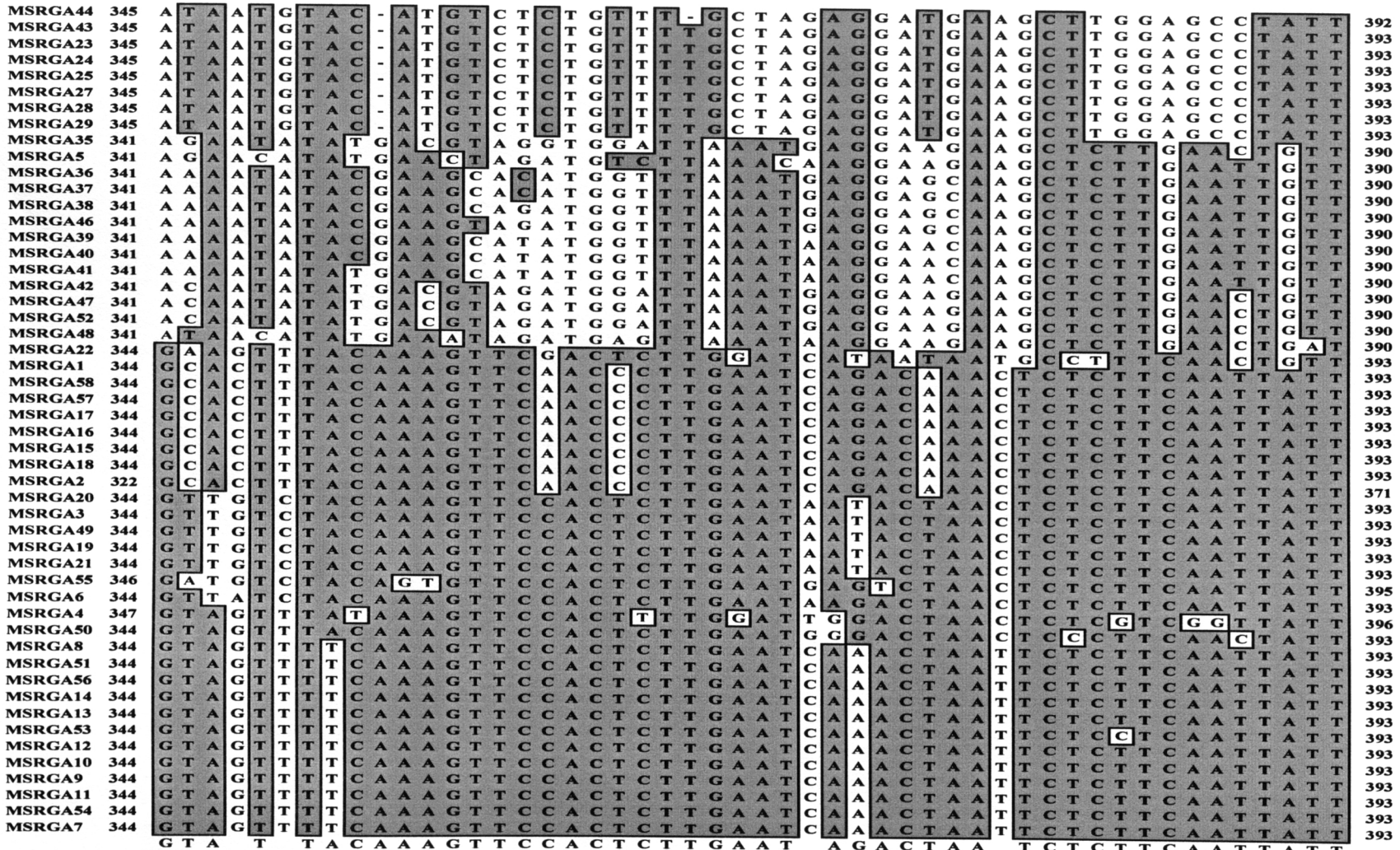


Figure 9d: Clustal W alignment of bases 146-195 of *Medicago sativa* resistance gene analogs


MSRGA44 196 GAAGGTAAAGGCCCGGCTTT


MSRGA44 246
MSRGA43 246
MSRGA23 246
MSRGA24 246
MSRGA25 246
MSRGA27 246
MSRGA28 246
MSRGA29 246
MSRGA35 242
MSRGA5 242
MSRGA36 242
MSRGA37 242
MSRGA38 242
MSRGA46 242
MSRGA39 242
MSRGA40 242
MSRGA41 242
MSRGA42 242
MSRGA47 242
MSRGA52 242
MSRGA48 242
MSRGA22 245
MSRGA1 245
MSRGA58 245
MSRGA57 245
MSRGA17 245
MSRGA16 245
MSRGA15 245
MSRGA18 245
MSRGA2 245
MSRGA20 245
MSRGA3 245
MSRGA49 245
MSRGA19 245
MSRGA21 245
MSRGA55 247
MSRGA6 245
MSRGA4 248
MSRGA50 245
MSRGA8 245
MSRGA51 245
MSRGA56 245
MSRGA14 245
MSRGA13 245
MSRGA53 245
MSRGA12 245
MSRGA10 245
MSRGA9 245
MSRGA11 245
MSRGA54 245
MSRGA7 245
295
295
295
295
295
295
295
295
290
290
290
290
290
290
290
290
290
290
290
290
290
293
293
293
293
293
293
293
293
293
293
293
293
293
293
293
293
293
293
293
293
293
293
293
293
293
293
293
293
271
293
293
293
293
293
293
295
293
293
293
293
293
293
293
293
293
293
293
293
293
293
293
293
293
293
293


MSRGA44 296
MSRGA43 296
MSRGA23 296
MSRGA24 296
MSRGA25 296
MSRGA27 296
MSRGA28 296
MSRGA29 296
MSRGA35 291
MSRGA5 291
MSRGA36 291
MSRGA37 291
MSRGA38 291
MSRGA46 291
MSRGA39 291
MSRGA40 291
MSRGA41 291
MSRGA42 291
MSRGA47 291
MSRGA52 291
MSRGA48 291
MSRGA22 294
MSRGA1 294
MSRGA58 294
MSRGA57 294
MSRGA17 294
MSRGA16 294
MSRGA15 294
MSRGA18 294
MSRGA2 272
MSRGA20 294
MSRGA3 294
MSRGA49 294
MSRGA19 294
MSRGA21 294
MSRGA55 296
MSRGA6 294
MSRGA4 297
MSRGA50 294
MSRGA8 294
MSRGA51 294
MSRGA56 294
MSRGA14 294
MSRGA13 294
MSRGA53 294
MSRGA12 294
MSRGA10 294
MSRGA9 294
MSRGA11 294
MSRGA54 294
MSRGA7 294
344
344
344
344
344
344
344
344
340
340
340
340
340
340
340
340
340
340
340
340
340
340
340
343
343
343
343
343
343
343
343
321
343
343
343
343
343
343
345
343
346
343
343
343
343
343
343
343
343


MSRGA44 345
MSRGA43 345
MSRGA23 345
MSRGA24 345
MSRGA25 345
MSRGA27 345
MSRGA28 345
MSRGA29 345
MSRGA35 341
MSRGA5 341
MSRGA36 341
MSRGA37 341
MSRGA38 341
MSRGA46 341
MSRGA39 341
MSRGA40 341
MSRGA41 341
MSRGA42 341
MSRGA47 341
MSRGA52 341
MSRGA48 341
MSRGA22 344
MSRGA1 344
MSRGA58 344
MSRGA57 344
MSRGA17 344
MSRGA16 344
MSRGA15 344
MSRGA18 344
MSRGA2 322
MSRGA20 344
MSRGA3 344
MSRGA49 344
MSRGA19 344
MSRGA21 344
MSRGA55 346
MSRGA6 344
MSRGA4 347
MSRGA50 344
MSRGA8 344
MSRGA51 344
MSRGA56 344
MSRGA14 344
MSRGA13 344
MSRGA53 344
MSRGA12 344
MSRGA10 344
MSRGA9 344
MSRGA11 344
MSRGA54 344
MSRGA7 344
392
393
393
393
393
393
393
393
390
390
390
390
390
390
390
390
390
390
390
390
390
390
393
393
393
393
393
393
393
393
371
393
393
393
393
393
395
393
396
393
393
393
393
393
393
393
393

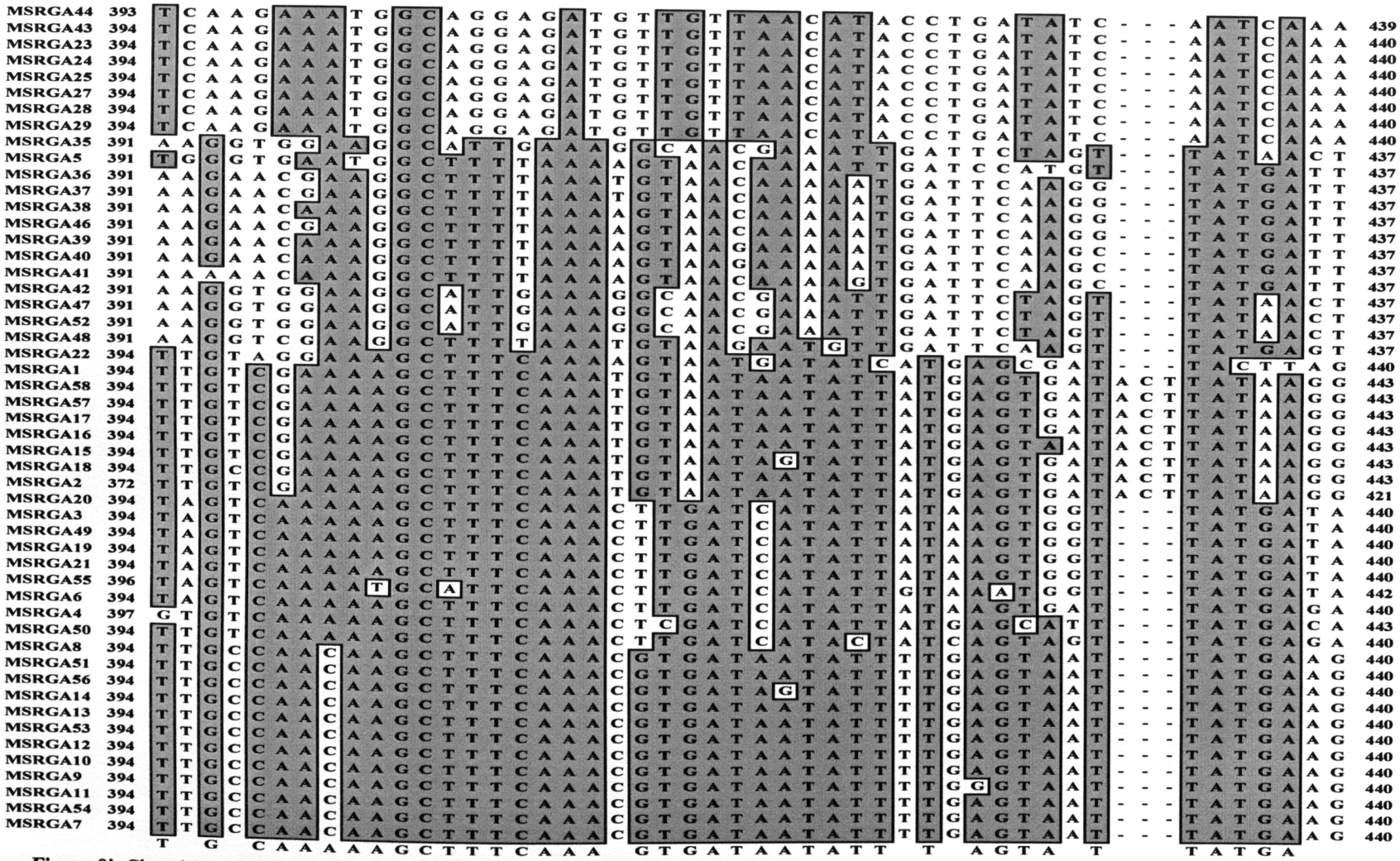


Figure 9i: Clustal W alignment of bases 394-443 of *Medicago sativa* resistance gene analogs.

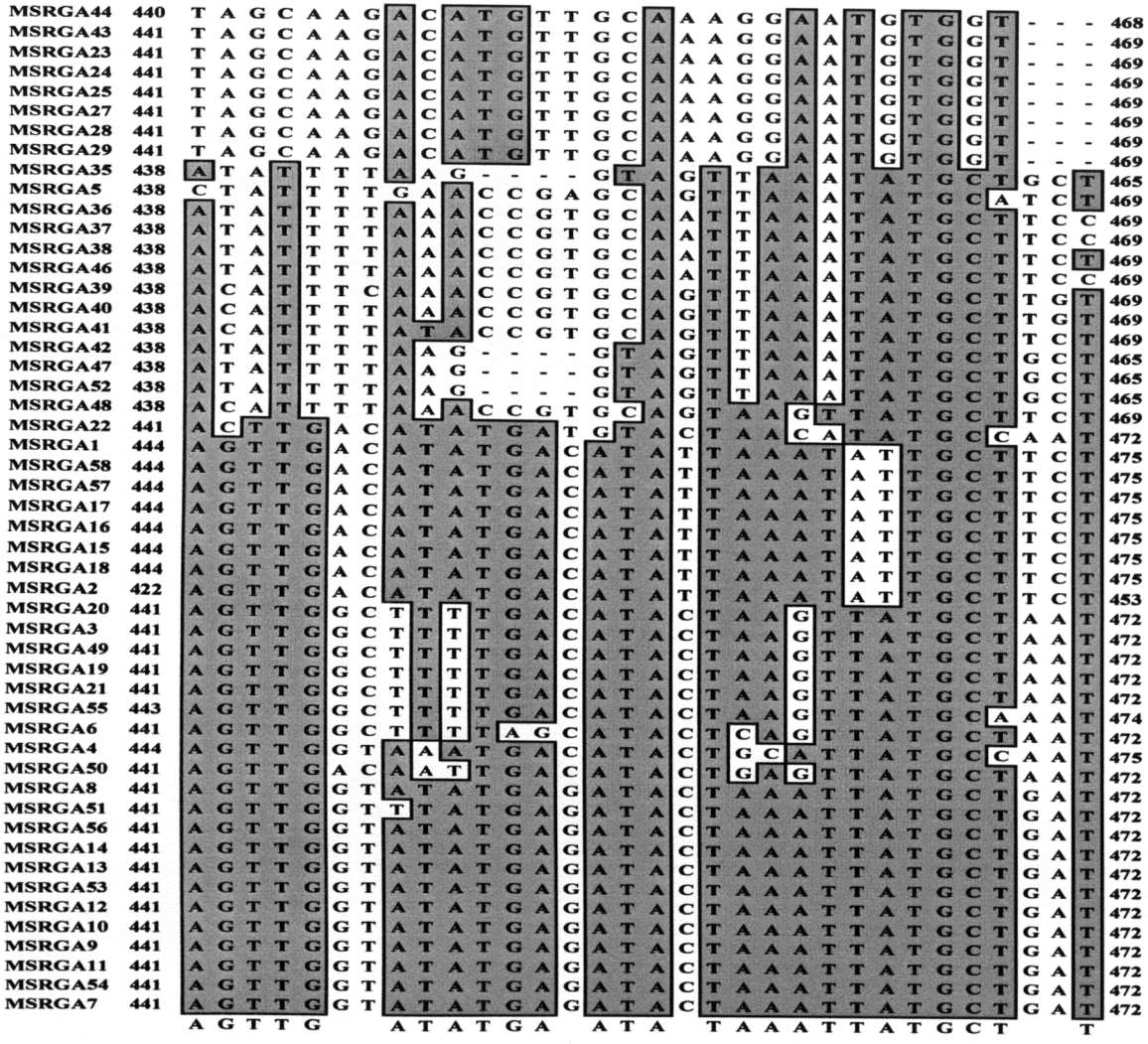


Figure 9j: Clustal W alignment of bases 444-475 of *Medicago sativa* resistance gene analogs.

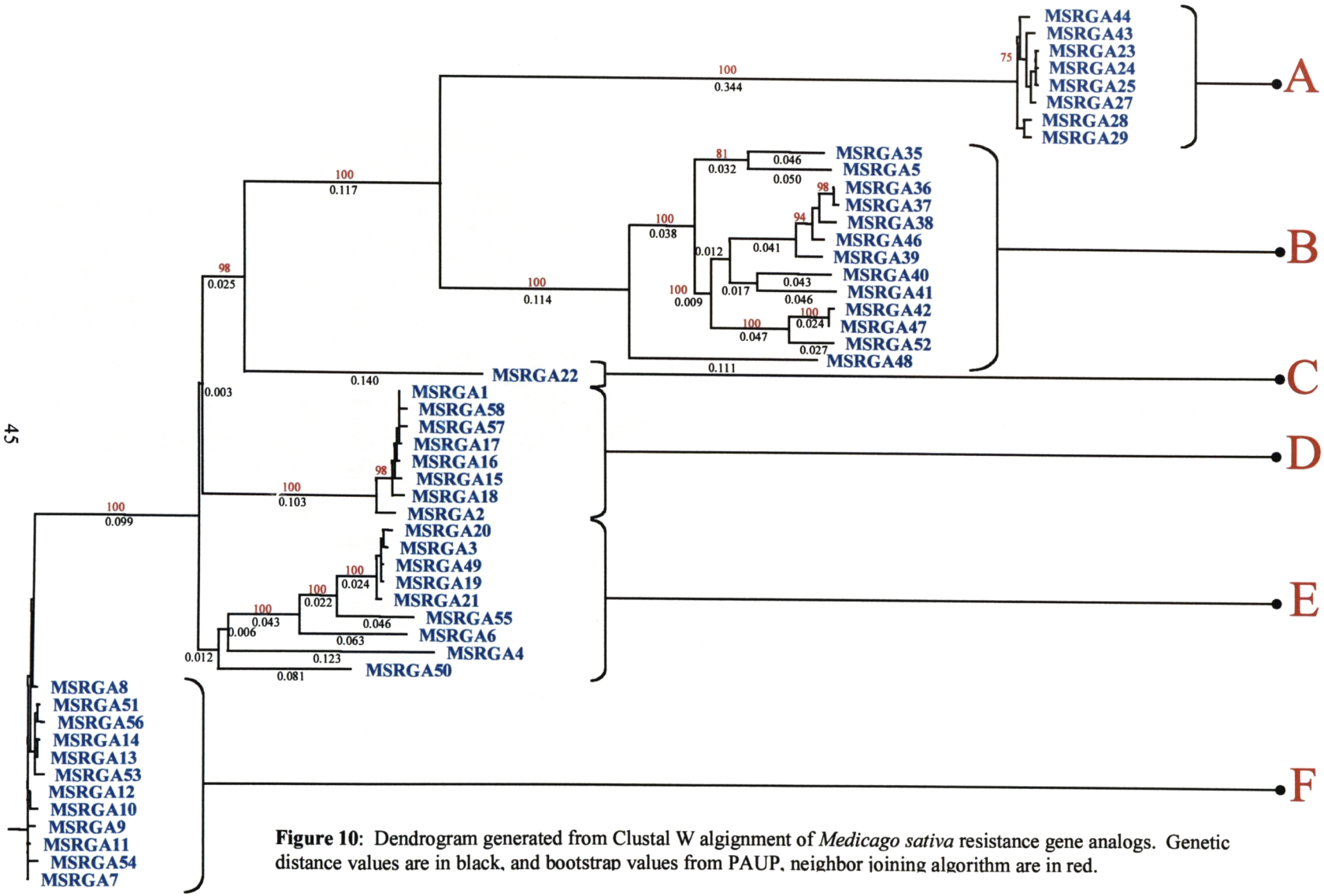


Figure 10: Dendrogram generated from Clustal W algignment of *Medicago sativa* resistance gene analogs. Genetic distance values are in black, and bootstrap values from PAUP, neighbor joining algorithm are in red.

Table 6: Percent homology between MSRGA families[a] at the nucleotide level.

Family		A		B								C	D	E				F
MSRGA		23	28	52	42	36	41	40	5	35	48	22	1	50	4	3	55	6	7
A	23	100	97	35	37	36	36	37	35	36	36	38	38	39	38	38	40	38	37
	28	97	100	36	37	37	37	38	36	37	37	39	38	39	39	38	40	38	38
B	52	35	36	100	94	87	86	81	79	86	75	51	48	48	47	47	49	46	51
	42	37	37	94	100	83	87	82	81	87	76	50	48	48	47	47	48	46	51
	36	36	37	87	83	100	88	85	81	80	75	51	49	49	48	48	49	47	51
	41	36	37	86	87	88	100	91	84	83	78	52	49	48	47	48	49	48	51
	40	37	38	81	82	85	91	100	89	87	78	50	48	49	47	48	48	49	52
	5	35	36	79	81	81	84	89	100	89	77	50	48	49	48	48	49	49	52
	35	36	37	86	87	80	83	87	89	100	76	50	49	49	48	47	48	48	52
	48	36	37	75	76	75	78	78	77	76	100	50	49	50	48	50	49	50	53
C	22	38	39	51	50	51	52	50	50	50	50	100	71	73	70	73	73	70	73
D	1	38	38	48	48	49	49	48	48	49	49	71	100	76	74	77	75	76	77
E	50	39	39	48	48	49	48	49	49	49	50	73	76	100	78	81	80	81	80
	4	38	39	47	47	48	47	47	48	48	48	70	74	78	100	71	76	77	75
	3	38	38	47	47	48	48	48	48	47	50	73	77	81	71	100	92	88	78
	55	40	40	49	48	49	49	48	49	48	49	73	75	80	76	92	100	86	76
	6	38	38	46	46	47	48	49	49	48	50	70	76	81	77	88	86	100	77
F	7	37	38	51	51	51	51	52	52	52	53	73	77	80	75	78	76	77	100

[a]See Figure 10.

MEDICAGO SATIVA RESISTANCE GENE ANALOG ENCODED SEQUENCES

The 51 sequences obtained were translated in order to look for the motifs which are characteristic of plant NBS regions (Meyers et al. 1999). Figure 11 shows the amino acid alignment of the most representative families for the MSRGA's as well as the closest NBS regions from the well-studied R genes. The encoded amino acid sequences showed the characteristic motifs of the nucleotide-binding site regions: kinase-2 and kinase-3a. The P-Loop and the GLPL motifs were part of the degenerate primer sequence. Sequence data showed that the inosine bases in the primers, which provided the degeneracy, were replaced by guanine bases in all of the clones (data not shown). However, the exact sequence is still unknown since the primers contained degeneracy and the specificity of any primer depends on the annealing conditions of the PCR. For this reason the primer sequences, corresponding to the P-Loop and the GLPL motifs were excluded from the sequence data analysis.

The encoded sequences were visually analyzed for the additional motifs found by Meyers et al. 1999 which are characteristic of NBS regions, and which separate plant NBS regions into the two major branches TIR-NBS and nonTIR-NBS (Table 3). A region which is identifiable by the similarity with the consensus sequence for the RNBS-C motif (Table 3) (Meyers et al. 1999) is found in all of the sequences obtained which confirms that they belong to the NBS-LRR super gene family (Figure 11c). Another motif, RNBS-A-nonTIR (Table3) found in the NBS region of resistance genes which do not contain the Toll-Interleukin Receptor domain (Meyers et al. 1999) was found in MSRGA23, MSRGA24, MSRGA25, MSRGA27, MSRGA28, MSRGA29, MSRGA43 and

MSRGA44 suggesting that these sequences correspond to NBS regions of nonTIR-NBS-LRR genes (Figure 11a). Meyers et al. 1999 also suggested that the last residue of the Kinase-2 domain could be used to predict with 95% accuracy whether an NBS region would fall in the TIR-NBS or the nonTIR-NBS phylogenetic trees. A tryptophan residue (W) was found at the end of the kinase-2 motif of the eight nonTIR-MSRGA's cloned which goes in accordance with the study done by Meyers et al. 1999 (Figure 11b).

Analogously, motifs similar to the RNBS-A-TIR (Table 3) were found in the rest of the sequences (Figure 11a) which suggests that they are NBS domains of resistance genes without TIR domains. None of the TIR-MSRGA's had the tryptophan residue in the kinase-2 motif, and instead they had either an aspartic acid (D) or an asparagine residue (N) (Figure 11b).

Of the 51 nucleotide-binding sites, all were found to contain the characteristic domains of plant resistance NBS. However, 8 of the 51 sequences were found to contain stop codons which distorted the amino acid sequences and consequently some of the motifs are not encoded (Figures 11c,d). The 8 clones that contain stop codons are: MSRGA2, MSRGA5, MSRGA44, MSRGA35, MSRGA42, MSRGA47, MSRGA52, MSRGA55 (See Table 5 and Figures 11c,d). MSRGA55 is likely to be a pseudo-gene since it contains multiple stop codons and consequently was unable to align with the other MSRGA's at the amino acid level. MSRGA2 contains a 22 bp deletion 5' of the kinase-3a motif which causes multiple stop codons 3' to the deletion. MSRGA5 sequences were found in 7 of the original 94 clones, and was found to encode a stop codon between the kinase-3a and the GLPL motifs. Likewise, MSRGA44, MSRGA35, MSRGA42, MSRGA47 and MSRGA52 encode a stop codon between the kinase-3a and

the GLPL motifs. Seven of the eight clones (All except MSRGA55) encode a methionine residue after the first stop codon, which suggests the start of translation of a different product. It is possible that these sequences which encode stop codons are pseudo-genes. However, the fact that the stop codons were found between the kinase-3a and the GLPL motifs (Figures 11c,d), the high frequency of MSRGA5 (Table 5), and the expression of MSRGA5 (discussed later) suggests that it is possible that they are truncated forms of the NBS-LRR regions. In such a case the three first motifs-P-Loop, Kinase-2, and Kinase-3a- would be sufficient to complete the function of the nucleotide binding site, and the LRR region would not be required for the function of the encoded gene.

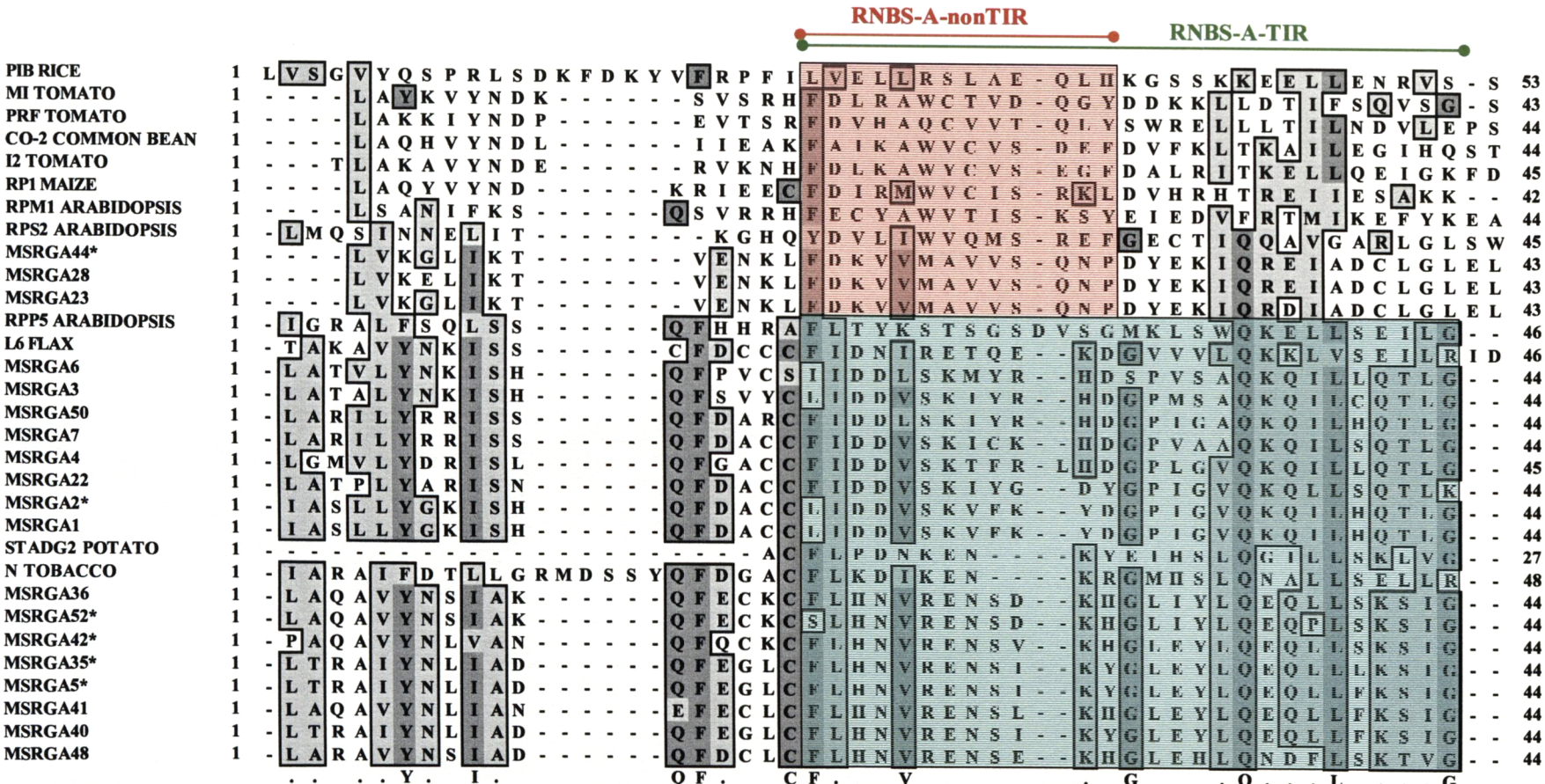


Figure 11a: Clustal W alignment of *Medicago sativa* resistance gene analog encoded sequences with nucleotide-binding site of characterized resistance genes from other species. Green shaded areas areTIR-NBS motifs, and red shaded areas are non-TIR-NBS motifs, according to Meyers et al. 1999 (See table 3).

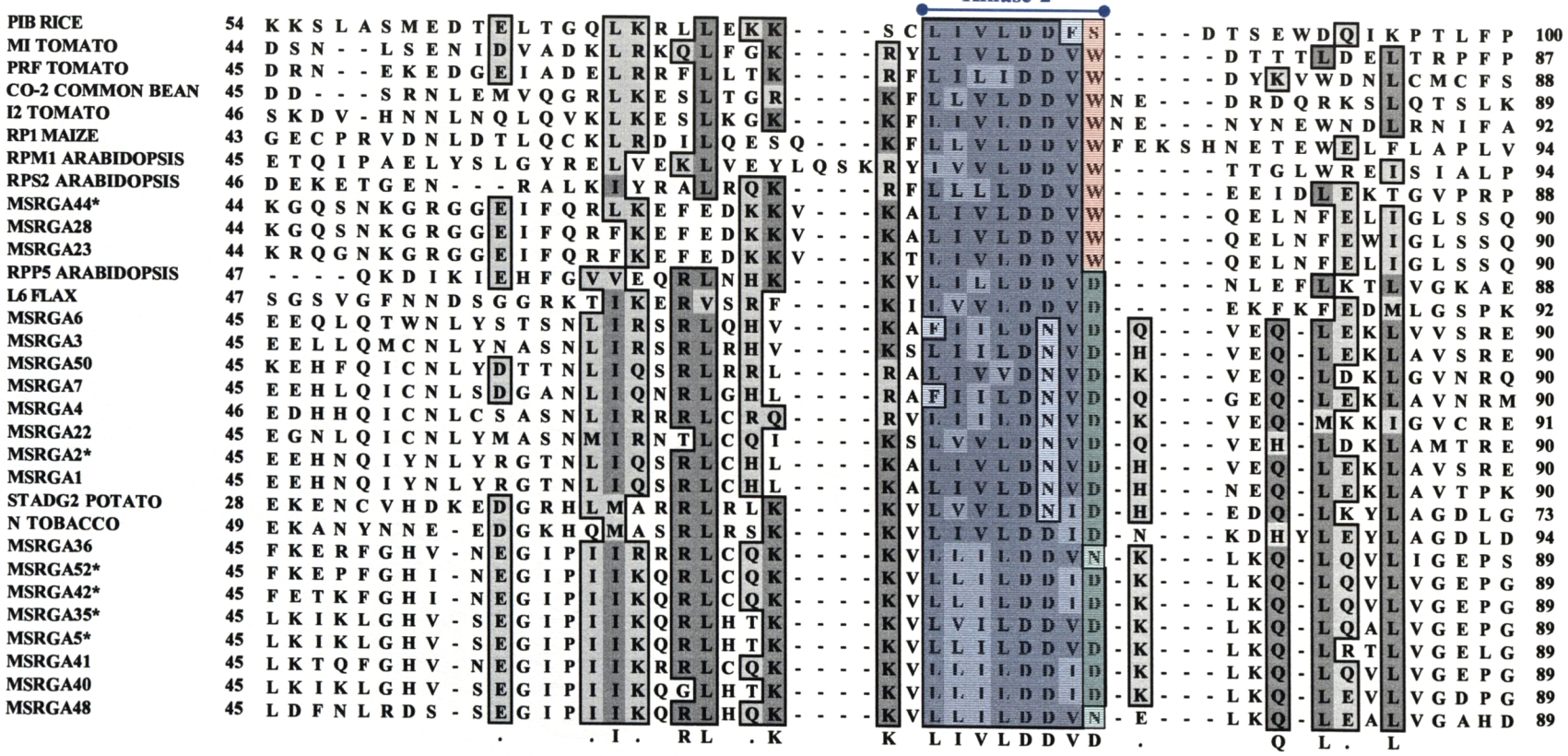


Figure 11b: Clustal W alignment of *Medicago sativa* resistance gene analog encoded sequences with nucleotide-binding site of characterized resistance genes from other species. Blue shaded areas represent motifs found in both types of NBS regions. Shown here is the Kinase-2 motif. The red shaded area is a Tryptophan residue characteristic of non-TIR-NBS regions, and the green shaded area is an aspartic acid or asparagine residue characteristic of TIR-NBS regions, according to Meyers et al. 1999 (See table 3).

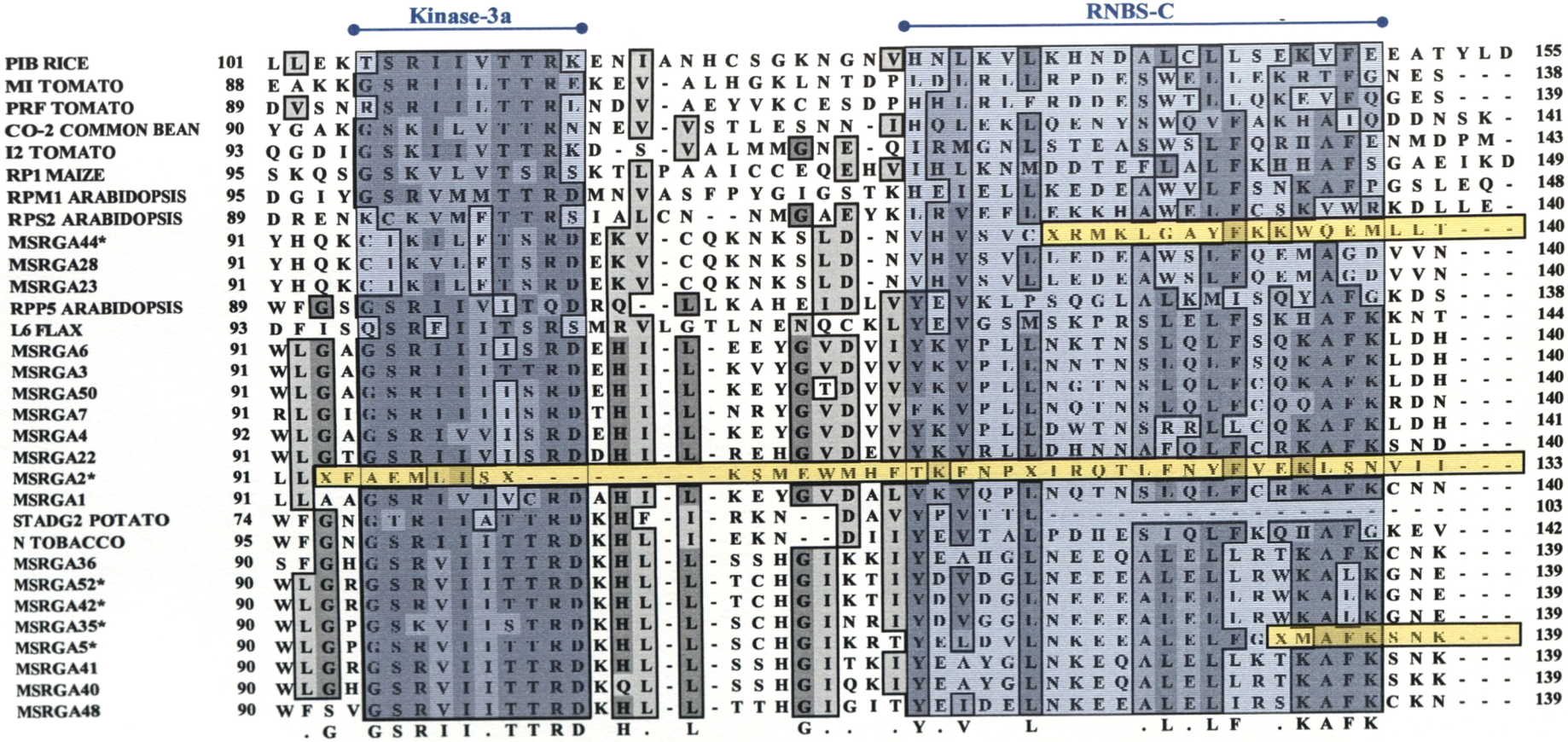


Figure 11c: Clustal W alignment of *Medicago sativa* resistance gene analog encoded sequences with nucleotide-binding site of characterized resistance genes from other species. Blue shaded areas represent motifs found in both types of NBS regions. Shown here are the Kinase-3a motif and the RNBS-C motif (Meyers et al. 1999) (Table 3). Yellow shaded areas mark sequences where stop codons were found. Yellow areas are marked starting from the first stop codon found.

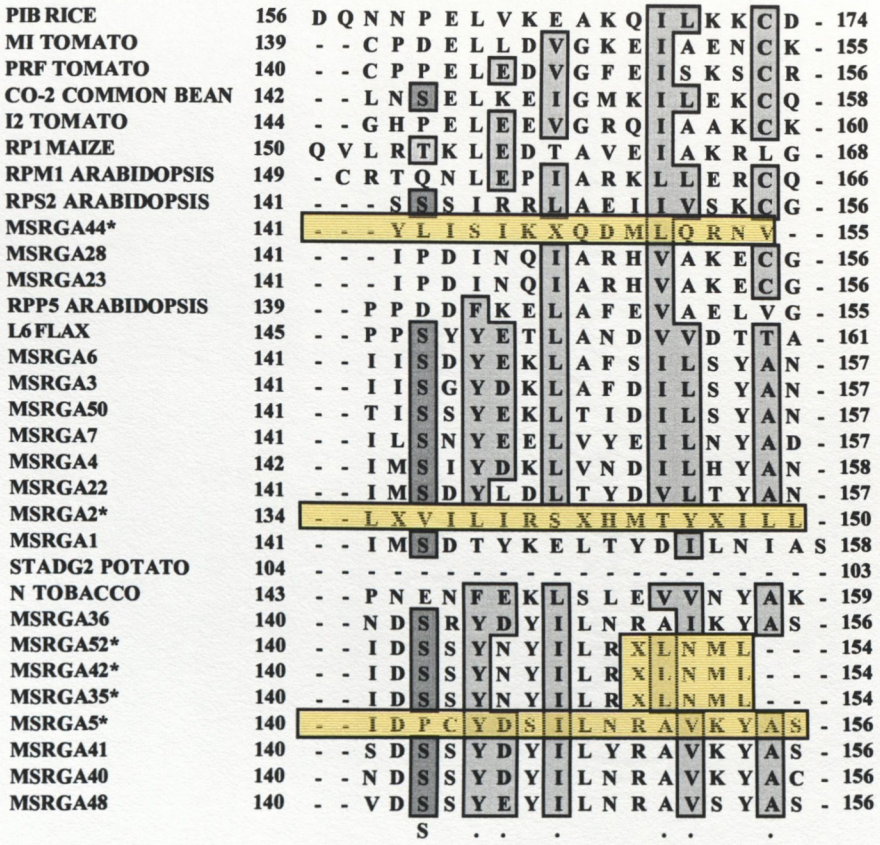


Figure 11d: Clustal W alignment of *Medicago sativa* resistance gene analog encoded sequences with nucleotide-binding site of characterized resistance genes from other species. Yellow shaded areas mark sequences where stop codons were found. Yellow areas are marked starting from the first stop codon

MEDICAGO SATIVA RESISTANCE GENE ANALOGS AND THE NUCLEOTIDE-BINDING SITE OF CHARACTERIZED DISEASE RESISTANCE GENES

The encoded sequences from the *Medicago sativa* clones confirmed that they are indeed resistance gene analogs. The cloned sequences were divided in six major clusters A, B, C, D, E and F. Additionally the 51 unique sequences are represented by eighteen families of MSRGA's that have at least 5% sequence divergence between them. The 18 families are a representative subsample of the variation in resistance gene analogs within alfalfa and were used to compare the variation within alfalfa MSRGA's with the NBS from characterized R genes.

Although MSRGA55 is considered as an MSRGA family it was excluded from this part of the study because the multiple stop codons interfere with the ability of ClustalW to align the sequences. For this reason, only 17 families were used to compare with the nucleotide-binding sites of R genes. Sequences which contain stop codons are indicated in the alignment in figure 11 a, b, c, d with an asterisk. These sequence were discussed earlier and should be excluded in the comparisons made here since the stop codons affect the encoded sequences, the protein alignments, and consequently the genetic distance trees generated by ClustalW and PAUP.

Figure 11 a, b, c, d shows the alignment generated by Clustal W of R gene NBS and MSRGA's. As discussed earlier, the encoded sequences showed the characteristic motifs of the NBS region. The alignment shows that the MSRGA's share homology with nucleotide-binding site regions of well-characterized R genes. The MSRGA's from alfalfa showed expected homology with R genes from other plants, which include

monocots and dicots: RPM1, RPS2 and RPP5 from Arabidopsis, L6 from flax, RP1 from maize, N from tobacco, PIB from rice, and PRF, MI, and I2 from tomato. STADG2 from potato is a tightly linked marker for virus resistance, and CO-2 is a candidate resistance gene for anthracnose resistance. CO-2 was included for comparison, since it represents molecular data from likely the first verified R gene from a legume species.

As expected, the highest homology between MSRGA's and R gene NBS's was at the highly conserved motifs: kinase-2 and kinase-3a. The kinase-2 motif was very highly conserved between all the R genes and MSRGA's compared. The kinase-3a motif; however, showed more variation than the kinase-2 motif. The kinase-3a motif in the nonTIR-MSRGA's differs from the expected consensus sequence: 5'-GSRIIITTRD-3' (Table 3) and instead shows: 5'-CIKILFTSRD-3'. The non-TIR MSRGA's (44,28,23) show a kinase-3a motif which differs from the other R genes and the other MSRGA's. However, by visual inspection the kinase-3a in these MSRGA's show more similarities with the kinase-3a motif found in the Arabidopsis RPS2 gene: 5'- KCKVMFTTRS-3' (Figure11c). The other NBS motifs (Table 3) discussed earlier were identifiable in the MSRGA's as well as the R gene NBS's. Those regions are highlighted in figures 11a, 11b and 11c.

In order to visualize the relative distance between MSRGA's and R genes, ClustalW generated a dendrogram based on the protein sequence alignment seen in figure 11 (Figure 12). The dendrogram showed that the MSRGA's and R genes cloned divided into two distinct branches: nonTIR-NBS and TIR-NBS. Two families MSRGA23 and MSRGA28 were placed in the nonTIR-NBS branch where other nonTIR-NBS-LRR genes are found (Figure 12).

The other fifteen families were located within the TIR-NBS branch where the other TIR-NBS-LRR genes are found including N from tobacco, L6 from flax, and RPP5 from Arabidopsis. Additionally, the TIR-NBS-MSRGA's are divided in two major groups. One group, which includes MSRGA36, MSRGA41, MSRGA40 and MSRGA48 as well as other families which contain stop codons (MSRGA52, MSRGA42, MSRGA35 and MSRGA5) all belong to cluster B (Figure 10). Cluster B represented in Table 7 by MSRGA48 placed closer to the virus resistance gene N from tobacco (Figure 12) showing 37% amino acid identity.

The other group, which includes MSRGA6, MSRGA3, MSRGA50, MSRGA7, MSRGA4, MSRGA22, and MSRGA1, contains sequences from clusters C, D, E and F (Figure10). For analysis purposes, one MSRGA family from each previously assigned cluster was used to compare the identity with known R genes (Table 7). MSRGA22 from cluster C and MSRGA7 from cluster F showed the highest identity (20% and 33%, respectively) with the tobacco N gene (Table 7). MSRGA1 from cluster D shared the highest amino acid identity (29%) with both the N gene and the L6 fungus resistance gene (Table 7). MSRGA 4 from cluster E showed the highest identity (29%) with the Arabidopsis fungus resistance gene RPP5.

Finally, MSRGA 23 which represents cluster A found within the nonTIR-NBS branch of the tree showed the closest identity (22%) with the candidate anthracnose resistance gene from common bean, CO-2. Although the homology values seem low between non-TIR-NBS MSRGA's and nonTIR-NBS genes, the R genes within this group also show low identity values. Identity between the studied nonTIR-NBS R genes only ranged between 13% to 49%.

In general the relation between the MSRGA's and the known R gene NBS regions does not show a clear pattern or an extremely high homology that might provide insight to the possible pathogen specificity of the cloned MSRGA's. However, the dendrogram (Figure 12), sequence data (Figures 11a-d), and the amino acid identities showed that two separate branches exist in alfalfa which are characterized by the presence or absence of a TIR domain at the amino terminal of the probable protein they encode. The TIR domain also has been found to have highly homologous motifs between TIR-NBS-LRR genes. This might allow the use of a PCR approach to amplify part of the gene found towards the N-terminus of the NBS region.

In addition, the TIR-NBS branch of the tree is divided in two major branches. Each branch contains approximately eight families of MSRGA's. One branch contains the MSRGA's that contain stop codons in frame (52,42,35,5) and other families that do not encode stop codons (36,41,40,48). The other branch contains seven families, none of which encode stop codons.

In summary, the dendrograms indicate that the MSRGA's found in alfalfa have three different evolutionary origins. One branch originates from non-TIR-NBS regions. The TIR-NBS part of the tree is also found to be divided. It is not clear whether this divergence is correlated with different mechanisms in which those genes are involved. The fact that stop codons were found mostly in one of the branches suggests that those genes have evolved a mechanism in which stop codons could have some function. Recent evidence from R genes has shown that stop codons play a role in alternative splicing (Dinesh-Kumar and Baker 2000). Additionally, P-Loop and UGA codon

containing genes are known to incorporate seleno-cysteine, which is considered the 21st amino acid.

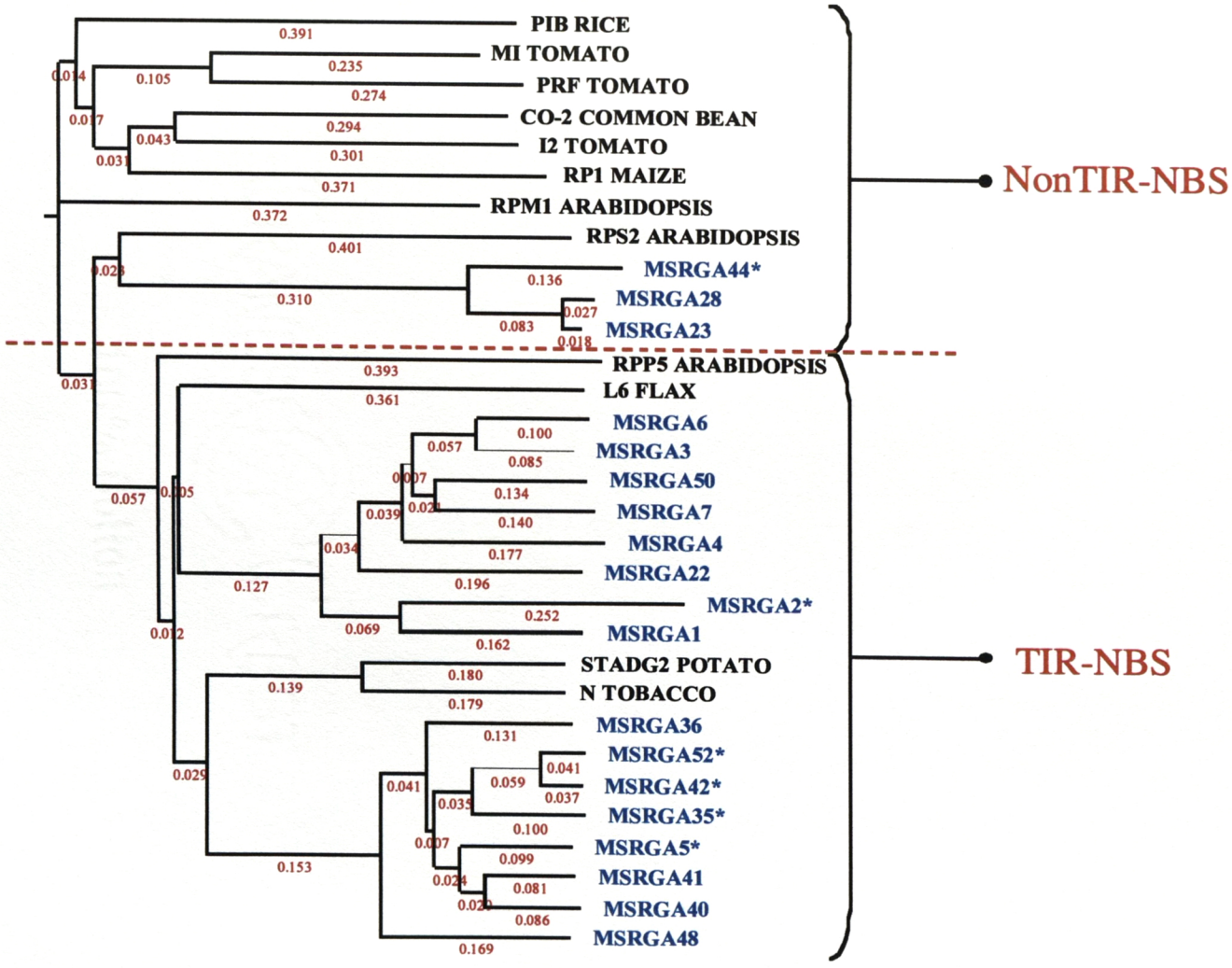


Figure 12: Dendrogram generated from Clustal W algignment of *Medicago sativa* resistance gene analogs (MSRGA's) and Nucleotide Binding Site region of disease resistance genes. Genetic distance values are in red . Tree divides in two branches: TIR-NBS and NonTIR-NBS. Additionally, the TIR-NBS branch subdivides in two major groups each containing families of MSRGA's.

Table 7: Percent homology between MSRGA clusters and with well studied R genes at the amino acid level.

Cluster		A	B	C	D	E	F	PIB	MI	PRF	CO-2	I2	RP1	RPM1	RPS2	N	RPP5	L6
MSRGA		23	48	22	1	4	7											
A	23	100	17	15	17	15	14	14	19	19	22	21	19	19	21	15	16	16
B	48	17	100	30	31	28	33	15	17	14	17	16	16	18	21	37	28	34
C	22	15	30	100	54	60	58	14	18	14	16	20	15	15	16	30	29	28
D	1	17	31	54	100	55	60	13	15	11	17	16	14	12	15	29	25	29
E	4	14	28	60	55	100	62	14	16	14	15	15	14	13	17	25	29	24
F	7	21	33	58	60	62	100	13	18	14	16	16	15	15	18	33	29	29
	PIB	14	15	14	13	14	13	100	18	20	18	20	15	16	13	16	17	13
	MI	19	17	18	15	16	18	18	100	49	31	33	23	24	17	20	20	14
	PRF	21	14	14	11	14	14	20	49	100	25	28	20	21	16	17	15	13
	CO-2	22	17	16	17	15	16	18	31	25	100	40	30	22	19	18	15	17
	I2	22	16	20	16	15	16	20	33	28	40	100	33	21	20	20	19	21
	RP1	19	16	15	14	14	15	15	25	20	30	33	100	20	19	17	15	17
	RPM1	19	18	15	12	13	15	16	24	21	22	21	20	100	18	16	15	18
	RPS2	21	21	16	15	17	18	13	17	16	19	20	19	18	100	18	12	17
	N	15	37	30	29	25	33	16	20	17	18	20	17	16	18	100	33	31
	RPP5	16	28	29	25	29	29	17	20	21	15	22	15	15	12	33	100	24
	L6	16	34	28	29	24	29	13	14	15	17	13	17	18	17	31	24	100

* Red labels indicate t the six clusters found by dendrogram analysis. On e MSRGA family from each cluster was used for comparison and is indicated in blue. The highest identity values for each MSRGA family is indicated in green.

ISOLATION OF EXPRESSED RESISTANCE GENE ANALOGS IN *MEDICAGO SATIVA*

The degenerate PCR primer approach showed positive results using genomic DNA. In order to determine whether these genes were expressed, total RNA was isolated from a single alfalfa plant, ISC35. The RNA was reverse transcribed to obtain the first strand cDNA, which was used as template for PCR. The primers used were the degenerate primers LM638 with REV7, and the non-degenerate primers FOR-NBS with REV-NBS. The sequences for the two sets of primers were the same except that the inosine bases that provided degeneracy were substituted with guanine bases.

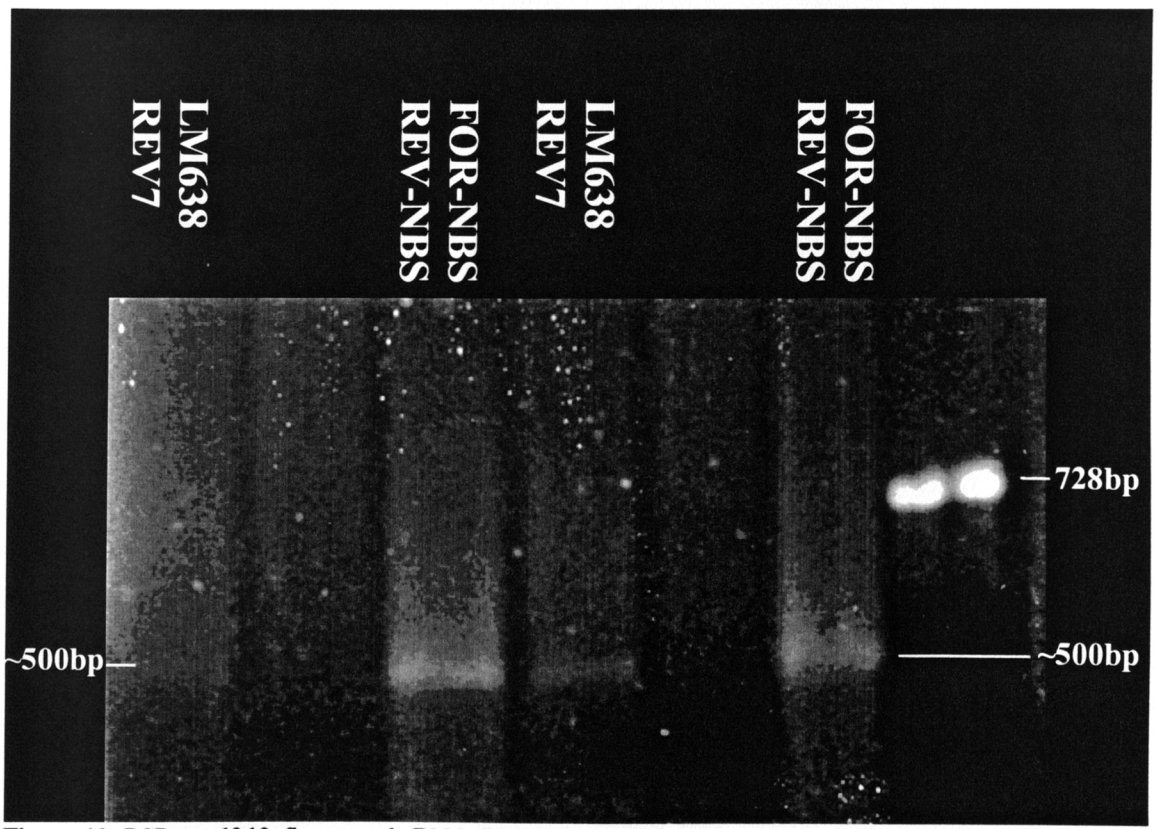


Figure 13: PCR on alfalfa first strand cDNA (ISC35) amplified a 500 bp product The first three lanes on the left are the same as the following three lanes to the right. NBS degenerate primers, LM638 and Rev7, were used in lanes 1 and 4. NBS non-degenerate primers, FOR-NBS and REV-NBS were used in lanes 3 and 6. In both cases, the non-degenerate PCR primers showed the best amplification and the amplification product in lane 3 was recovered for cloning.

PCR with the NBS primers (degenerate and non-degenerate) showed amplification of a 500 bp product (Figure 13). The band in lane 3 showed the best amplification and was cut out of the gel and cloned in the pGEM vector. Many white colonies were obtained out of which 48 clones were selected; the DNA extracted and cryo-preserved.

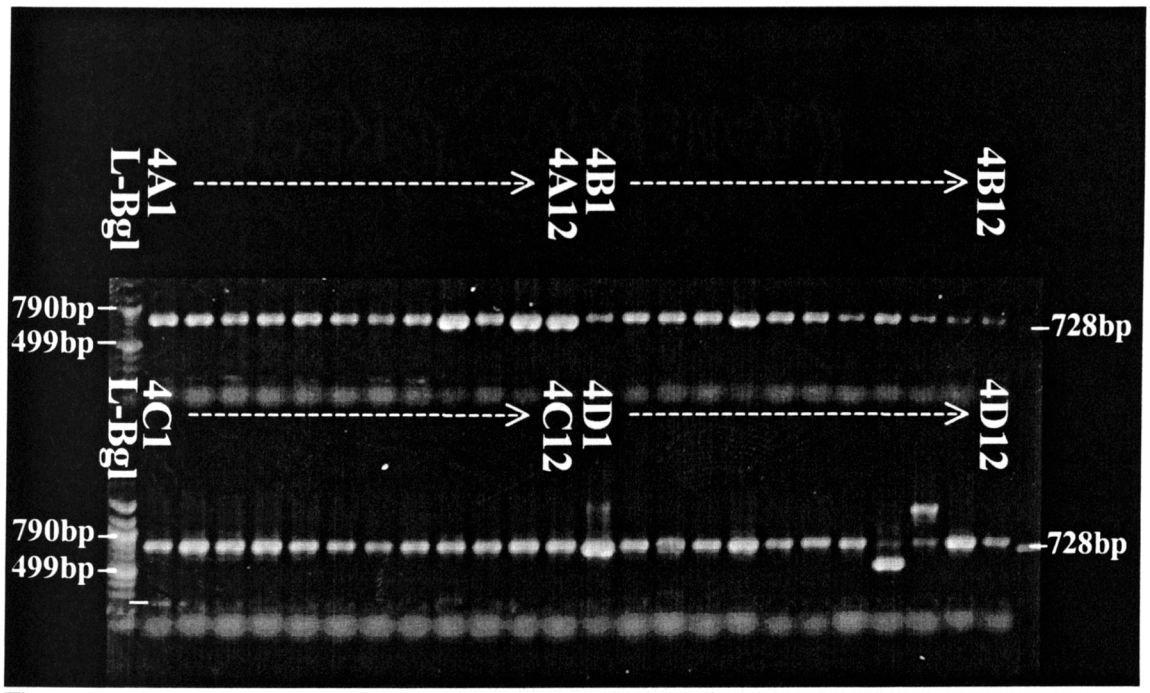


Figure 14: Forty-eight clones were confirmed by PCR with M13 primers.

The 48 clones were PCR amplified using M13 primers. The cDNA clones showed a uniform and expected 740 bp product that corresponds to the 505-512 bp NBS product plus approximately 245 bp from the cloning vector (Figure 14). Eleven clones of the 48 were selected in order to look for variation among the clones. The 11 clones were digested with Tsp509I. The digest showed that there was indeed variation within the cDNA clones obtained (Figure 15).

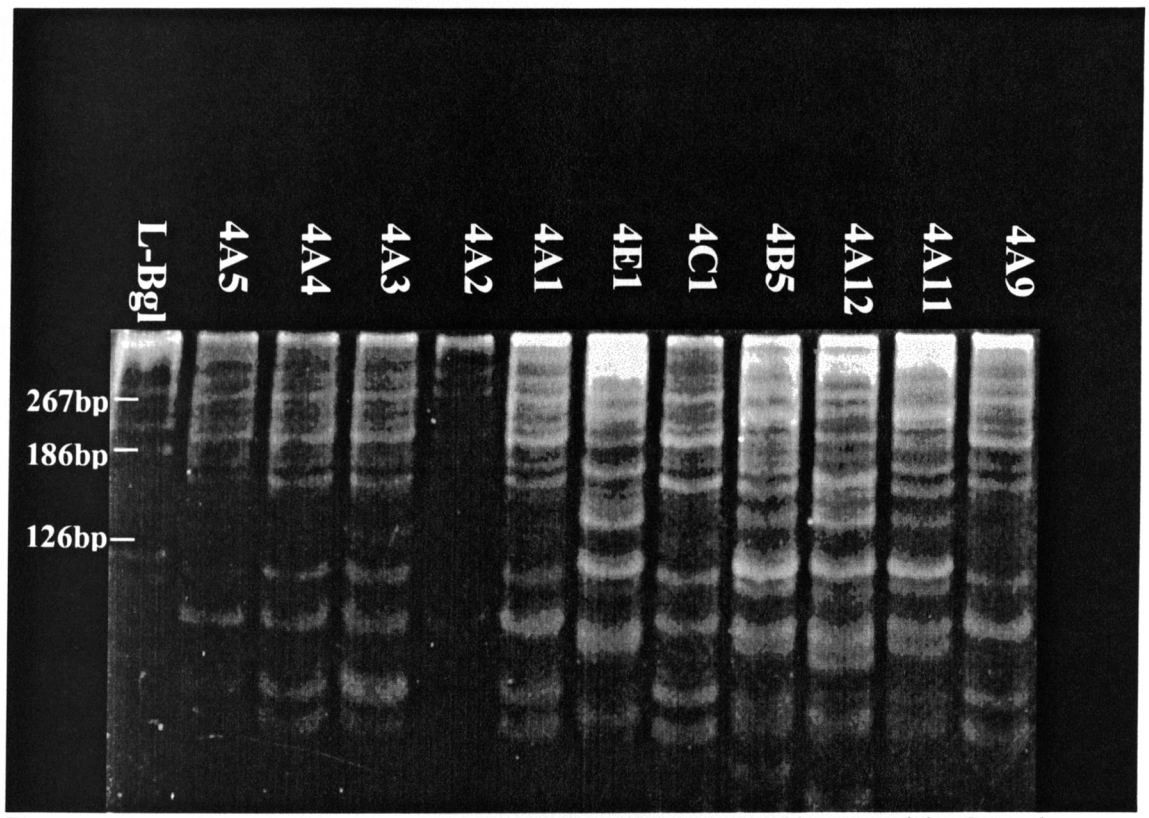


Figure 15: Individual cDNA clones showed some variation between. For this reason eight cDNA clones were selected for sequencing.

Eight cDNA clones were then selected for sequencing: 4A9, 4A11, 4A12, 4B5, 4C4, 4C7 4D1 and 4D4. Reliable sequence data were obtained from five of the clones. All five clones showed high homology to previously sequenced MSRGA's from genomic DNA. Clones 4A9, 4A11, and 4B5 showed highest homology to MSRGA1 (>99%). Clone 4D1 showed the highest homology to MSRGA5 (>99%) and clone 4A12 showed the highest homology to MSRGA41 (>99%). This confirms that at least those three families of MSRGA's are being transcribed in *Medicago sativa*; however, it is possible that most or all of the other families are being transcribed as well.

MSRGA5, which was determined earlier to have a stop codon in the region between the kinase-3a and the GLPL domains, was also found within the cDNA clones. As mentioned, clone 4D1 showed 99% homology with MSRGA5. The region where the

stop codon is found in MSRGA5 is 100% homologous to the same region in clone 4D1: 5'-GGG TGA ATG GCT-3' which encodes NH2- G opal M A -COOH. This confirms that the transcription product indeed has a stop codon at that position, but it is not known whether that product is actually translated. However, opal codons have been found in selenocysteine proteins where a UGA codon is translated and is in charge of incorporating a selenocysteine residue (Guimaraes et al. 1996). Additionally, some of these proteins contain a Walker-A or P-Loop like motif. This similarity might provide further information about the possible functioning of some R genes.

Another alternative for the possible role of stop codons in R genes is that truncated forms and alternative splicing of the R gene encoded products play a role in resistance. This recently has been confirmed in studies using the N gene in tobacco, where scientists found that the ratio of two altenatively spliced messages obtained from cDNA, the Nl and the Ns, played a role in N gene-mediated resistance. The Nl message is a truncated form of the N gene product caused by a UGA codon in the reading frame (Dinesh-Kumar and Baker 2000).

RESISTANCE GENE ANALOG DIVERSITY WITHIN THE GENUS *MEDICAGO*

The nucleotide binding site family of resistance gene analogs was found to be very diverse in *Medicago sativa*, and the PCR approach we used to study that diversity gave very satisfactory results. This allowed us to use a similar approach to assess the variation in resistance gene analogs within the genus *Medicago*. The LM638 and REV7 degenerate primers were used on 13 different species from the genus *Medicago: M. lupulina, M. polymorpha, M. orbicularis, M. rigidula, M. rugosa, M. murex, M. ruthenica, M. medicaginoides, M. platycarpa, M. praecox, M. radiata, M. blancheana, M. constricta and M. sativa* (UC123 and Riley 117A). Degenerate PCR gave good amplification of a clear 500 bp which is the expected size for NBS regions.

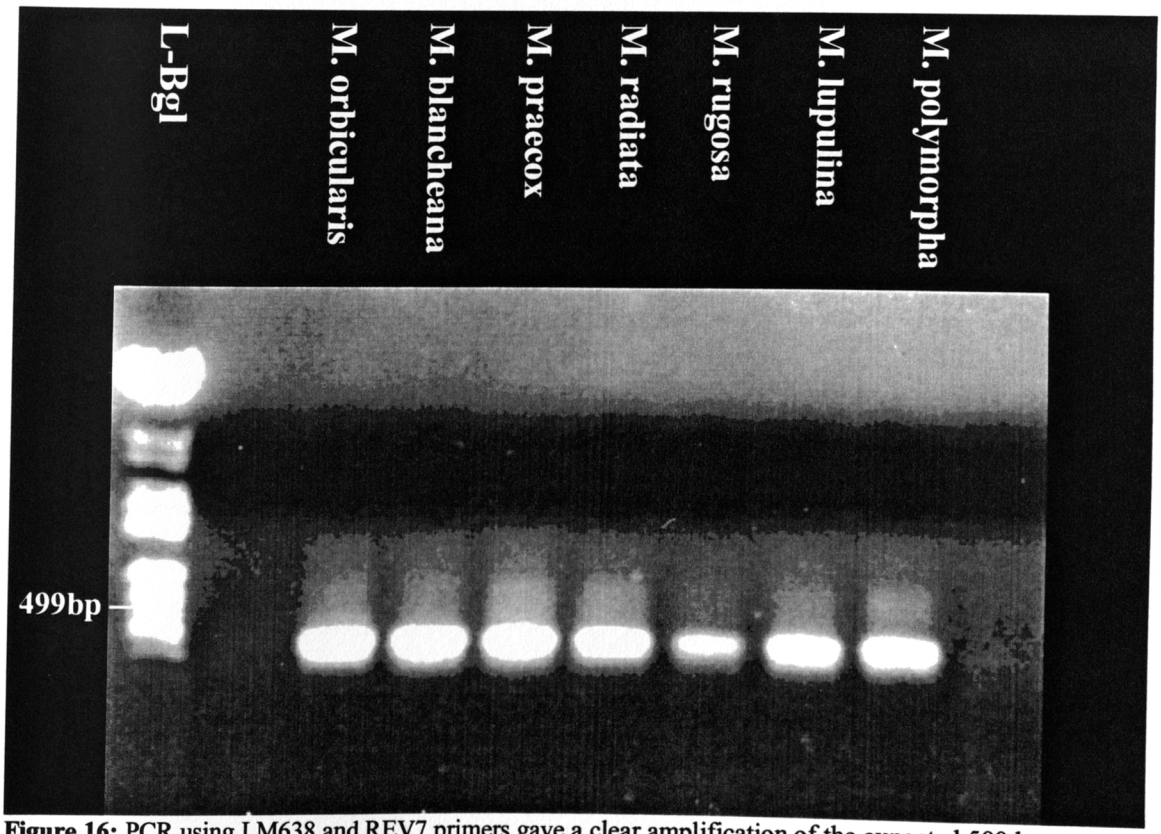


Figure 16: PCR using LM638 and REV7 primers gave a clear amplification of the expected 500 bp product in several species of the genus *Medicago*.

The degenerate PCR products were digested with Tsp509I in order to study the resistance gene analog diversity between the species mentioned.

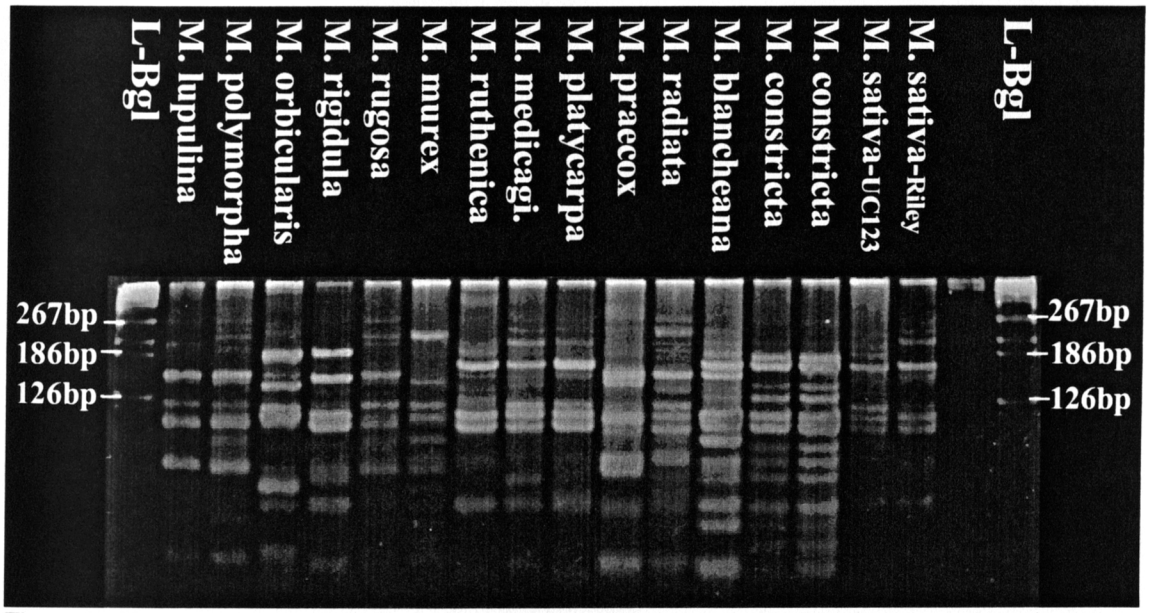


Figure 17: Restriction digests with Tsp509I of the clear 500 bp product from the degenerate PCR showed variation in the banding patterns from the different *Medicago* species.

Restriction digests showed that there is significant variation among resistance gene analogs in the genus *Medicago* (Figure 17). The variation in RGA's within the genus represents a pool of candidate R genes that might be useful in the future for introducing disease resistance from distantly related wild species.

Wild relatives in the genus *Medicago* have been suggested previously as a potential gene pool for disease and pest resistance genes. Several species of the genus have been screened for disease and pest resistance. Very high levels of resistance against races 1 and 2 of anthracnose have been reported in *M. littoralis, M. murex, M. rigidula, M. tenoreana, M. truncatula* and high levels in *M. arborea.* Resistance to spring blackstem (*Phoma medicaginis*) in M. dzhwakhetica and M. suffruticosa; to bacterial wilt

(*Corynebacterium insidiosum*) in *M. arborea*, and to *Stemphylium* leafspot in *M. cancellata* have also been reported (Reviewed by Quiros and Bauchan 1988).

Attempts to cross plants from the *Medicago sativa* complex with wild *Medicago* species have failed due to several barriers that make the species cross incompatible. Among these are ploidy level, chromosome rearrangement, pollen morphology and poor seed development (Reviewed by Quiros and Bauchan 1988).

The use of molecular techniques for the isolation of disease and pest resistance genes as well as other techniques such as plant transformation might enable breeders to over-come the barriers between the wild species of the genus *Medicago* and the cultivated *Medicago sativa*. This would enable alfalfa breeders to make use of the R gene pool that we have found to exist within the genus *Medicago* in order to incoporate it in breeding programs.

CONCLUSIONS

The present study provided a characterization of the nucleotide-binding site family of resistance gene analogs in *Medicago sativa*. We found that the MSRGA family is highly diverse in alfalfa and within the genus *Medicago*. In alfalfa we found 51 unique sequences that encoded the characteristic motifs of the nucleotide-binding sites of disease and pest resistance genes. The 51 unique sequences were found divided in 6 major clusters according to dendrogram and bootstrap analysis. Some clusters showed little variation between the sequences (<5% nucleotide sequence divergence) in the cluster. Sequences that showed less than 5% sequence divergence might represent alleles of homologous resistance genes. Sequences that diverged by more than 5% were considered as members of different families. The 51 sequences were classified into eighteen families of *Medicago sativa* resistance gene analogs. Three of the 18 families were confirmed to be transcribed. Additionally, families in *Medicago sativa* suggest three different evolutionary origins by the occurrence of three groups in dendrogram analysis: two groups with TIR-NBS characteristics, and one with nonTIR-NBS characteristics.

Our present study showed that there is significant variation in resistance gene analogs in the genus *Medicago* that represents a diverse pool of resistance genes. The sequences provided will be a useful starting point for the isolation of complete functional resistance genes from *Medicago sativa* and related species.

The characterization of the MSRGA family will also contribute in finding molecular markers for use in marker-assisted selection. The use of techniques such as CAPS, used successfully in potato, will enable us to look for polymorphisms within domains of

candidate resistance genes that consequently would provide tightly linked disease resistance markers. The identification of markers within functional gene domains associated with resistance will likely increase the efficiency of marker assisted selection in autotetraploid species like alfalfa.

As research increases involving genetic analyses in *Medicago sativa* and *Medicago truncatula* the R gene diversity that we describe here in alfalfa and in the genus *Medicago* will gain importance in the development of disease and pest resistant alfalfa cultivars in the future.

In summary, the MSRGA family in *Medicago sativa* and in the genus *Medicago* provides a rich source of naturally occurring genetic diversity with the potential to confer strong resistance against a wide range of pathogens and pests. Any strategy successful in incorporating resistance from this diverse gene pool into elite cultivars will reduce the application of agricultural chemicals, and will increase the efficiency of food production.

The research efforts provided by this study look towards the improvement of agriculture through the pursuit of scientific knowledge with beneficial implications to the preservation of the environment, and the well being of agriculturally dependent societies.

REFERENCES

Aarts N, Metz M, Holub E, Staskawicz BJ, Daniels MJ, Parker JE (1998) Different requirements for EDS1 and NDR1 by disease resistance genes define at least two R-gene mediated signaling pathways Proc Natl Acad Sci USA 95: 10306-10311

Anderson PA, Lawrence GJ, Morrish BC, Ayliffe MA, Finnegan EJ, Ellis JG (1997) Inactivation of the flax rust resistance gene M associated with loss of a repeated unit within the leucine-rich repeat coding region. Plant Cell 9: 641-651

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

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

Bourne HR, Sanders DA, McCormick F (1991) The GTPase superfamily: conserved structure and molecular mechanism. Nature 349: 117-127

Botella MA, Parkers JE, Forst LN, Bittner-Eddy PD, Beynon JL, Daniels MJ, Holub EB, Jones JDG (1998) Three genes of the Arabidopsis Rpp1 complex resistance locus recognize distinct *Peronospora parasitica* avirulence determinants. Plant Cell 10:1847-1860.

Brummer EC, Bouton, JH, Kochert G (1993) Development of an RFLP map in diploid alfalfa. Theor Appl Genet 86: 329-332

Buschges R, Hollricher K, Panstruga R, Simons G, Wolter M, Fritjers A, van Daelen R, van der Lee T, Diergaarde P, Groenendijk J, Topsch S, Vos P, Salamini F, Schulze-Lefert P (1997) The barley Mlo gene: a novel control element of plant pathogen resistance.

Cai D, Kleine M, Kifle S, Harloff HJ, Sandal NN, Marcker KA, Klein-Lankhorst RM, Salentijn EMJ, Lange W, Stiekema WJ, Wyss U, Grundler FMW, Jung C (1997) Positional cloning of a gene for nematode resistance in sugar beet. Science 275: 832-834

Collins N, Drake J, Ayliffe M, Sun Q, ellis J, Hulbert S, Pryor T (1999) Molecular characterization of the maize Rp1- rust resistance haplotype and its mutants. Plant Cell 11: 1365-1376

Collins NC, Webb CA, Seah S, Ellis JG, Hulbert SH, Pryor A (1998) The isolation and mapping of disease resistance gene analogs in maize. Mol Plant-Microbe Interact 11: 968-978

Creusot F, Macadré C, Cana EF, Riou C, Geffroy V, Sévignac M, Dron M, Langin T (1999) Cloning and molecular characterization of the NBS-LRR subfamily located in the vicinity of the Co-2 locus for anthracnose resistance in *Phaseolus vulgaris.* Genome 42: 254-264

Dellaporta S, Woods J, Hicks J (1983) A plant DNA minipreparation, version II. Plant Mol Biol Rep 1:19-21

Dinesh-Kumar SP, Baker BJ (2000) Alternatively spliced N resistance gene transcripts: their possible role in tobacco mosaic virus resistance. Proc Natl Acad Sci USA 97: 1908-1913

Dixon MS, Jones DA, Keddie JS, Thomas CM, Harrison K, Jones JD (1996) The tomato Cf-2 disease resistance locus comprises two functional genes encoding leucine-rich repeat proteins.

Ellis J, Jones D (1998) Structure and function of proteins controlling strain-specific pathogen resistance in plants. Curr Opin Plant 2: 273-279

Flor HH (1956) The complementary genic systems in flax and flax rust. Adv Genet 8: 29-54

Grant MR, Godiard L, Straube E, Ashfield T, Lewald J, Sattler A, Innes RW, Dangl JL (1996) Structure of the Arabidopsis RPM1 gene enabling dual specificity disease resistance. Science 269:843-846

Guimaraes MJ, Peterson D, Vicari A, Cockes BG, Copeland NG, Gilbert DJ, Jenkins NA, Ferrick DA, Kastelein RA, Bazan JF (1996) Identification of a novel selD homolog from eucaryotes, bacteria, and archaea: is there an autoregulatory mechanism in selenocysteine metabolism? Proc Natl Acad Sci USA 93:15086-15091

Hill RR, Shenk JS, Barnes RF (1988) Breeding for yield and quality. In: Hanson AA, Barnes DK, Hill RR (eds) Alfalfa and alfalfa improvement. ASA, Madison, Wisconsin pp 809-825

Johal and Briggs (1992) Reductase activity encoded by the HM1 disease resistance gene in maize. Science 258: 985-987

Jones DA, Thomas CM, Hammond-Kosack KE, Balint-Kurti PJ, Jones JD (1994) Isolation of the tomato Cf-9 gene for resistance to *Cladosporium fulvum* by transposon tagging. Science 266: 789-793

Kanazin V, Marek LF, Shoemaker RC (1996) Resistance gene analogs are conserved and clustered in soybean. Proc Natl Acad Sci USA 93: 11746-11750

Lawrence GJ, Finnegan EJ, Ayliffe MA, Ellis JG (1995) The L6 gene for flax resistance is related to the arabidopsis gene RPS2 and the tobacco viral resistance gene N. Plant Cell 7: 1195:1206

Leath KT, Erwin DC, Griffin GD (1988). Disease and Nematodes. In: Hanson AA, Barnes DK, Hill RR (eds) Alfalfa and alfalfa improvement. ASA, Madison, Wisconsin pp 621-670

Leister D, Ballvors A, Slamini F, Gebhardt C (1996) A PCR-based approach for isolating pathogen resistance genes from potato with potential for wide application in plants. Nature Genetics 14:421-429

Manglitz GR, Ratcliffe RH (1988) Insects and Mites. In: Hanson AA, Barnes DK, Hill RR (eds) Alfalfa and alfalfa improvement. ASA, Madison, Wisconsin pp 671-704

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

Martin GB (1999) Functioinal analysis of plant disease resistance genes and their downstream effectors. Curr Opin Plant 2: 273-279

McDowell JM, Dhandaydham M, Long TA, Aarts MGM, Goff S, Holub EB, Dangl JL (1998) Intragenic recombination and diversifying selection contribute to the evolutions of downy mildew resistance at the Rpp8 locus of Arabidopsis. Plant Cell 10: 1861-1874.

Meyers BC, Dickerman AW, Michelmore RW, Sivaramakrishnan S, Sobral BW, Young ND (1999) plant disease resistance genes encode members of an ancient and diverse protein family within the nucleotide-binding superfamily. Plant J 20: 317-332

Milligan SB, Bodeau J, Yaghoobi J, Kaloshian I, Zabel P, Williamson VM (1998) The root Knot nematode resistance gene Mi from tomato is a member of the leucine zipper nucleotide-binding, leucine-rich repeat family of plant genes. Plant Cell 10: 1307-1319

Nam YW, Penmetsa RV, Endre G, Uribe P, Kim D, Cook DR (1999) Construction of a bacterial artificial chromosome library of Medicago truncatula and identification of clones containing ethylene-response genes. Theor Appl Genet 98:638-646

National Aricultural Statistics Service Agricultural Statistics Board. United States Department of Agriculture. (1999) http://www.usda.gov/nass.

Ori N, Eshed Y, Paran I, Presting G, Aviv D, Tanksley S, Zamir D, Fluhr R (1997) The I2C family from the wilt disease resistance locus I2 belongs to the nucleotide binding, leucine-rich repeat superfamily of plant resistance genes. Plant Cell 9: 521-532

Parker JE, Coleman MJ, Szabo V, Frost LN, Schmidt R, van der Biezen EA, Moores T, Dean C, Daniels MJ, Jones JD The arabidopsis downy mildew resistance gene RPP5 share similarity to the toll and interleukin-1 receptors with N and L6. Plant Cell 9: 879-894

Quiros CF, Bauchan GR (1988) The genus *Medicago* and the origin of the *Medicago sativa* complex. In: Hanson AA, Barnes DK, Hill RR (eds) Alfalfa and alfalfa improvement. ASA, Madison, Wisconsin pp 93-124

Rossi M, Goggin FL, Milligan SB, Kaloshian I, Ullman DE, Williamson VM (1998) The nematode resistance gene Mi of tomato confers resistance against the potato aphid. Proc Natl Acad Sci USA 95: 9750-9754

Salmeron JM, Oldroyd GED, Tommens CMT, Scofield SR, Kim H, Lavelle DT, Dahlbeck D, Staskawicz BJ (1996) Tomato Prf is a member of the leucine-rich repeat class of plant disease resistancegenes and lies embedded within the Pto kinase gene cluster. Cell 86: 123-133

Saraste M, Sibbald PR, Wittinghofer A (1990) The P-loop a common motif in ATP- and GTP-binding proteins. Trends Biochem Sci 15: 430-434

Shen KA, Meyers BC, Islam-Faridi MN, Chin DB, Stelly DM, Michelmore RW (1998) Resistance gene candidates identified by PCR with degenerate primers map to clusters of resistance genes in lettuce. Mol Plant-Microbe Interact 8: 815-823

Song WY, Wang GL, Chen LL, Kim HS, Pi LY, Holsten T, Gardner J, Wang B, Zhai WX, Zhu LH, et al. (1995) A receptor kinase-like protein encoded by the rice disease resistance gene, Xa21. Science 270: 1804-1806

Sorri VA, Watanabe KN, Valkonen JPT (1999) Predicted kinase-3a motif of a resistance gene analogue as a unique marker for virus resistance. Theor Appl Genet 99: 164-170

Staskawicz BJ, Ausubel FM, Baker BJ, Ellis JG, Jones JDG (1995) Molecular genetics of plant disease resistance. Science 268: 661-667

Traut TW (1994) The functions and consensus motifs of nine types of peptide segments that form different types of nucleotide-binding sites. Eur J Biochem 222: 9-19

Walker JE, Saraste M, Runswick MJ, Gay NJ (1982) Distantly related sequences in the α- and β- subunits of ATP synthetase, myosin, kinases and other ATP-requiring enzymes and a common nucleotide-binding fold. EMBO J 1: 945-951

Wang Z, Yano M, Yamanouchi U, Iwamoto M, Monna L, Hayasaka H, Katayose Y, Sasaki T (1999) The Pib gene for rice blast resistance belongs to the nucleotide

binding and leucine-rich repeat class of plant disease resistance genes. Plant J 19: 55-64

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

Yoshimura S, Yamanouchi U, Katayose Y, Toki S, Wang Z, Kono I, Kurata N, Masahiro Y, Iwata N, Sasaki T (1998) Expression of Xa1, a bacterial blight-resistance gene in rice, is induced by bacterial inoculation. Proc Natl Acad Sci USA 95: 1663-1668

Yu YG, Buss GR, Maroof MAS (1996) Isolation of a superfamily of candidate disease-resistance genes in soybean based on a conserved nucleodtide-binding site. Proc Natl Acad Sci USA 93:11751-11756