Phylogeny and Genomic Organization of the TIR and Non-TIR NBS-LRR Resistance Gene Family in *Medicago truncatula*

Hongyan Zhu,^{1, 2} Steven B. Cannon,³ Nevin D. Young,³ and Douglas R. Cook¹

¹Department of Plant Pathology, University of California, Davis 95616, U.S.A.; ²Graduate Program in Genetics, Texas A&M University, College Station 77843, U.S.A.; ³Department of Plant Pathology and Department of Plant Biology, University of Minnesota, St. Paul 55108, U.S.A.

Submitted 6 December 2001. Accepted 12 February 2002.

Sequences homologous to the nucleotide binding site (NBS) domain of NBS-leucine-rich repeat (LRR) resistance genes were retrieved from the model legume *M. truncatula* through several methods. Phylogenetic analysis classified these sequences into TIR (toll and interleukin-1 receptor) and non-TIR NBS subfamilies and further subclassified them into several well-defined clades within each subfamily. Comparison of *M. truncatula* NBS sequences with those from several closely related legumes, including members of the tribes Trifoleae, Viceae, and Phaseoleae, reveals that most clades contain sequences from multiple legume species. Moreover, sequences from species within the closely related Trifoleae and Viceae tribes (e.g., *Medicago* and *Pisum* spp.) tended to be cophyletic and distinct from sequences of *Phaseoleae* species (e.g., soybean and bean). These results suggest that the origin of major clades within the NBS-LRR family predate radiation of these Papilionoid legumes, while continued diversification of these sequences mirrors speciation within this legume subfamily. Detailed genetic and physical mapping of both TIR and non-TIR NBS sequences in *M. truncatula* reveals that most NBS sequences are organized into clusters, and few, if any, clusters contain both TIR and non-TIR sequences. Examples were found, however, of physical clusters that contain sequences from distinct phylogenetic clades within the TIR or non-TIR subfamilies. Comparative mapping reveals several blocks of resistance gene loci that are syntenic between *M. truncatula* and soybean and between *M. truncatula* and pea.

Disease resistance (R) genes have been cloned from numerous plant species, and these genes confer resistance to a wide range of plant pathogens, including bacteria, fungi, oomycetes, viruses, and nematodes (Baker et al. 1997; Bent 1996; Hammond-Kosack and Jones 1997). In addition to R genes encoding serine-threonine kinase, leucine-rich repeat (LRR) domains, or both (e.g., *Pto*, *Xa21*, and the *Hcr* family of R genes), the majority of R genes encode proteins with a centrally located nucleotide binding site (NBS) and a C-terminal LRR region (Hammond-Kosack and Jones 1997). The NBS-LRR family of R genes can be further subdivided into two subfamilies based on the presence or absence of a region homologous to the toll and interleukin-1 receptor (TIR) domain at their

N-terminus (Meyers et al. 1999; Pan et al. 2000a; Young 2000), namely TIR and non-TIR. While the non-TIR NBS-LRR subfamily does not encode a TIR domain, most members of this subfamily instead encode a putative coiled-coil (CC) domain in their N-terminus (Pan et al. 2000a).

TIR and non-TIR NBS-LRR R genes can also be distinguished by amino acid motifs found within the NBS domain itself (Meyers et al. 1999; Pan et al. 2000a; Young 2000). Thus, phylogenetic analysis of NBS sequences can effectively differentiate the TIR and non-TIR subfamilies, and subfamily-specific polymerase chain reaction (PCR) primers have been used for selective amplification of NBS sequences from either one of the two subfamilies (Pan et al. 2000a).

The TIR and non-TIR NBS-LRR R genes also appear distinct at the functional level, based on their putative involvement in different signal transduction pathways (Aarts et al. 1998). The inferred functional separation is based on work in *Arabidopsis* in which R genes from the TIR and non-TIR subfamilies have been found to operate through distinct signaling pathways requiring either *EDS1* (Parker et al. 1996) or *NDR1* (Century et al. 1997), respectively. These studies suggest that the N-terminal TIR or CC domains or the corresponding NBS motifs, or both, play an important role in the bifurcation of signaling pathways leading to disease resistance. Interestingly, both database searches and experimental procedures reveal that the non-TIR subfamily is widely distributed in both monocot and dicot species, whereas the TIR subfamily appears to be restricted to dicot species (Meyers et al. 1999; Pan et al. 2000a).

A large number of NBS sequences with homology to R genes, so-called “resistance gene analogs” (RGAs), have been isolated from many plant species through PCR amplification with degenerate primers (Kanazin et al. 1996; Leister et al. 1996; Yu et al. 1996), as well as by genome and cDNA sequencing projects. NBS sequences are abundant in plant genomes (Meyers et al. 1999). For example, analysis of the whole genome sequence of *Arabidopsis* reveals 150 NBS-encoding genes, including 85 TIR NBS-LRR genes, 43 non-TIR NBS-LRR genes, and numerous truncated TIR- or CC-NBS genes that lack an LRR domain (The Arabidopsis Genome Initiative 2000). While precise functions have not been established for each of these RGAs, many of them map to clusters of functionally defined R gene loci (Kanazin et al. 1996; Leister et al. 1996; Yu et al. 1996). Thus, RGAs are valuable as potential sources of active R genes.

R genes and RGAs are typically clustered in plant genomes (The Arabidopsis Genome Initiative 2000; Michelmore and Meyers 1998). This local redundancy of R gene sequences may

Corresponding author: Douglas R. Cook; Telephone: +1.530.754.6561; Fax: +1.530.754.6617; E-mail: drcook@ucdavis.edu

provide a reservoir of genetic variation from which new specificities can evolve (Dangl and Holub 1997; Ellis et al. 2000; Lynch and Conery 2000; Michelmore and Meyers 1998; Young 2000). In fact, individual clusters may confer specific resistance to a wide range of different pathogens and pathogen genotypes (van der Vossen et al. 2000). Although R gene clusters frequently contain closely related sequences, the individual genes may exhibit closer phylogenetic relationships to sequences from other clusters within the same genome (Kanazin et al. 1996; Leister et al. 1998; Meyers et al. 1998, 1999; Pan et al. 2000b; Wei et al. 1999). Several mechanisms including duplication, unequal crossing-over, gene conversion, diversifying selection, and transposable elements have been proposed to contribute to the structure of R gene clusters and the evolution

of resistance specificities (Ellis et al. 2000; Michelmore and Meyers 1998; Richter and Ronald 2000; Young 2000).

Medicago truncatula (barrel medic) is an annual, self-fertile, diploid legume species that has become a model for studying various aspects of legume genomics and biology (Cook 1999). The suitability of *M. truncatula* as a nodal species for legume genomics will depend, in part, on its utility for answering questions about the function and evolution of legume genes and genomes, including the large and malleable group of resistance genes. Here, we report the cloning, sequencing, and detailed genetic and physical mapping of NBS sequences that are representative of the NBS-LRR superfamily in *M. truncatula*. We examine the genomic organization of distinct RGA phylogenetic groups in the *M. truncatula* genome and compare the se-

Fig. 1. Phylogenetic tree of *Medicago truncatula* nucleotide binding site (NBS) sequences. Numbers under the branches represent the percentage of 1,000 bootstrap replications supporting the particular nodes, and only those with >70% supported are labeled. TIR indicates the branch containing the toll and interleukin-1 receptor (TIR) type NBS sequences, and non-TIR indicates the branch comprised of the non-TIR type NBS sequences. T1 to T4 are four major clades within the TIR subfamily; N1 to N8 are eight clades within the non-TIR subfamily. Within each defined clade, sequences are generally within 1.0 distance unit of one another and have at least 70% bootstrap support. Asterisks indicate the sequences that were used as probes for BAC library screening. To facilitate comparison, the resistance gene expressed sequence tags are shown in red.

Fig. 2. Phylogenetic tree of nucleotide binding site (NBS) sequences from *Medicago truncatula* and several other legume species. The prefix indicates the species from which the NBS sequences were derived. Mt = *Medicago truncatula*, Gm = *Glycine max*, Ms = *Medicago sativa*, Ps = *Pisum sativum*, Pv = *Phaseolus vulgaris*, Ca = *Cicer arietinum*, and Ce = *Cajanus cajan*. The Pv sequences were from Rivkin and associates (1999). These sequences were further classified into numerous clades (a to m). Within each defined clade, sequences are generally within 1.0 distance unit of one another and have at least 70% bootstrap support. Green indicates species from gallegoid, and red indicates species from the tribe Phaseoleae.

quence diversity found in *M. truncatula* with that observed in other legumes. Features of mapped RGAs show both similarities and differences with R-gene organization in other species, helping to delimit the extent of generalizations that can be drawn about mechanisms of evolution in this important gene family.

RESULTS

PCR amplification and database searching of NBS sequences from *M. truncatula*.

With the intent of characterizing a broad set of sequences with homology to the NBS domain of plant resistance genes, we used a combination of PCR with degenerate primers, hybridization with RGAs from related legumes, and database queries. For PCR, two pairs of degenerate primers, designated F1-R1 and F2-R2, were used to amplify *M. truncatula* genomic DNA. Primers F1 and R1 were designed to anneal to the P-loop and the GLPL amino-acid motifs of RGAs, respectively, and amplified fragments of approximately 500 bp. The primer pair F2 and R2 was designed for targeted isolation of non-TIR NBS sequences only and yielded PCR products of roughly 700 bp. About 100 clones from each PCR primer pair were sequenced. Searches of these sequences with the BLASTX algorithm against the National Center for Biotechnology Information database (Altschul et al. 1997) revealed significant similarity to known resistance genes or RGAs. In most cases, the highest-scoring BLASTX hits were RGAs from other closely related legumes (data not shown). Preliminary alignment of DNA sequences permitted their separation into eight TIR and fourteen non-TIR NBS homology groups. Each homology group was composed of sequences exhibiting at least 90% identity. One representative from each of these homology groups was selected for further analysis. The first pair of primers yielded only TIR-type NBS sequences. This is somewhat surprising, considering that both TIR and non-TIR sequences contain the conserved P-loop and GLPL motifs. The same degenerate primers also amplified largely TIR-type NBS sequences from soybean (*Glycine max*) (Kanazin et al. 1996).

The bias may be due, in part, to specific amino acid residues or biased codon usage within the conserved motifs of TIR or non-TIR subfamilies among different species. The second primer pair generated only non-TIR NBS sequences.

A search of the The Institute for Genomic Research (TIGR) *M. truncatula* GeneIndex (MtGI) using 14 known NBS-LRR R-gene protein sequences as queries yielded 147 unique R-gene-like expressed sequence tags (R-ESTs) with E values less than e^{-20} . These 147 sequences were represented by 54 tentative consensus sequences (TCs) and 93 singleton ESTs. The 147 unique sequences can be further grouped into 107 TIR and TIR-related NBS sequences and 40 non-TIR NBS sequences. This ratio is very close to that found in *Arabidopsis* sp. (The Arabidopsis Genome Initiative 2000).

Phylogenetic analysis of NBS sequences from *M. truncatula*.

M. truncatula sequences (51) were translated to amino acid sequences and used in a phylogenetic analysis. The 51 sequences were composed of 22 sequences from PCR amplification, 28 from the MtGI database, and 1 from BAC-end sequences. TCs or singleton ESTs (28) were chosen because they contain the complete sequence between the P-loop and GLPL motifs. A neighbor-joining tree (Fig. 1) constructed from the aligned sequences unambiguously grouped these sequences into two major branches, comprising TIR and non-TIR subfamilies. The two branches were further supported by the presence or absence of characteristic subfamily-specific consensus sequences, consistent with the results reported by Meyers and associates (1999) and Pan and associates (2000a).

The distribution of RGAs within Figure 1 reveals a high level of divergence of *M. truncatula* sequences within both the TIR and non-TIR subfamilies. Within the TIR subfamily, the sequences can be subdivided into four well-supported major clades, designated as T1, T2, T3, and T4. The sequence identities within a clade range from 50 to 98%, while sequence identities among clades are less than 30%. Similarly, the non-TIR NBS sequences also fall within several strongly supported and typically well-represented clades, designated N1 to N8. Sequences within a clade generally show 50 to 95% identities. Interestingly, clades N5 to N8 contain numerous EST sequences that are distantly related to the non-TIR NBS sequences from clades N1 to N4 (<20% similarity), and this distance is almost equivalent to that found between TIR and non-TIR NBS sequences from *M. truncatula*. Nevertheless, their placement within the non-TIR subfamily was supported by the presence of characteristic amino acid residues within the NBS domain, the pre-P-loop sequences, as well as the comparison with their significant non-TIR homologs from *Arabidopsis* strains (CAB38788, BAB08976, and BAB08632) for which complete gene sequences are known. The classification of N5 to N8 as non-TIR was further supported by a larger scale phylogenetic analysis involving sequences from other species (described below), in which these ESTs were unambiguously assigned to the non-TIR subfamily. These results support the findings of Meyers and associates (1999) and Pan and associates (2000a) that R genes show evidence of ancient divergence into separate subfamilies followed by significant diversification within each subfamily.

For both the TIR and non-TIR subfamilies, R-ESTs and PCR-derived sequences occurred together in most clades (Fig. 1), indicating that the wide diversity of genomic sequences in *M. truncatula* is mirrored by a similar complexity of expressed genes. Interestingly, clades N5 to N8 and T3 and T4 were represented only by ESTs in *M. truncatula* although, for several of these sequences, close homologs were identified in other legume species (discussed below).

Fig. 3. Southern blot results for representative nucleotide binding site (NBS)-containing BAC contigs. **A, B, C, and D,** represent four contigs containing TIR-nucleotide binding site sequences. Asterisks indicate the BACs that have been mapped. Arrows indicate BAC clone vectors. Contigs A and B are closely linked on linkage group (LG) 4, and contigs C and D are closely linked on LG 6. The identity of DNA probes used for Southern blot analyses are given in Table 1.

Comparative phylogenetic analysis of NBS sequences between *Medicago truncatula* and other legumes.

For purposes of comparing the diversity of *M. truncatula* RGA sequences with that of other legumes, we compiled the set of RGA sequences that have been characterized from six other Papilionoid legume species. These species included three species from the "gallegoid" clade, which also contains *M. truncatula* (i.e., *Pisum sativum*, *Medicago sativa*, and *Cicer arietinum*) and three species from within the phylogenetically distinct tribe Phaseoleae (i.e., *G. max*, *Phaseolus vulgaris*, and *Cajanus cajan*). All sequences were either PCR-derived or ESTs retrieved from the NCBI nonredundant database. Sequences were chosen to represent the range of distinct sequence classes that have been identified within the given species. Several general relationships are evident from the derived phylogenetic tree shown in Figure 2. First, all of the major clades within the TIR subfamily and most major clades within the non-TIR subfamily are composed of sequences from two or more species. Second, *M. truncatula* sequences were identified within each clade with the exception of TIR clade d, and thus, the diversity of RGAs in *M. truncatula* is representative of the diversity of known RGAs from the six other legume species included in Figure 2.

In some cases, minor clades appear to represent RGA lineages that arose (or were lost) uniquely in certain taxa, while in other cases the appearance of species-specific minor clades is likely due to biased or incomplete sampling. For example, clade j (Fig. 2), an apparently *M. truncatula*-specific clade, contains numerous *M. truncatula* sequences represented by Mt02a, Mt08a, and Mt23. Using any of these clones as probes of soybean BAC library filters under permissive hybridization conditions produced no positives. However, using Mt03a from clade k, which contains no soybean sequences, as a probe of the same soybean BAC library identified more than 100 positive BACs (data not shown). Thus, methodical sampling of the *M. truncatula* genome served to uncover additional sequences within the Phaseoleae tribe.

In many clades, the most similar sequences tended to originate from closely related species. For example, in the TIR subfamily, sequences from *M. truncatula* are frequently closest to sequences from *M. sativa* (with sequence identity as high as 95%), followed at greater distances to sequences from *Pisum sativum*, *G. max*, and *Phaseolus vulgaris* (e.g., clade a and c,

Fig. 2). The observation that sequences of *M. truncatula* and *M. sativa* frequently occur together in minor clades indicates a similar diversity of TIR sequences within the two species. Similar results were obtained for the non-TIR subfamily, in which sequences of *M. truncatula* and *Pisum sativum* (e.g., clades h, k, and l) and *G. max* and *Phaseolus vulgaris* (clades h and m) were often coincident in minor clades. These findings correlate well with the established phylogenetic relationships between these legume species (Doyle 1998). In the case of the non-TIR subfamily, comparison between *M. truncatula* and *M. sativa* sequences was not possible, as *M. sativa* has not been systematically sampled for this RGA subfamily.

Physical characterization of NBS sequences at the level of individual BAC contigs.

To determine the physical organization of RGAs within the *M. truncatula* genome, a *M. truncatula* BAC library (Nam et al. 1999) was screened by hybridization with representative NBS sequences as probes. The probes were selected to represent most of the well-defined clades (Fig. 1, sequences marked with an asterisk), and thus, are expected to represent the majority of NBS sequences in the *M. truncatula* genome. Positive BAC clones were assembled into contigs by restriction enzyme fingerprinting (data not shown), and the copy number and diversity of RGA sequences were assessed by Southern blot analysis as shown by example in Figure 3 and as listed in Table 1.

For the TIR subfamily, eight probes from the T1 and T2 clades gave a total of 89 positive BAC clones, of which 68 were analyzed by fingerprinting and assembled into 17 contigs. Southern blot analysis of these 17 contigs, using the same eight probes, gave a total of 53 nonredundant hybridization bands, providing a minimum estimate of copy number within clades T1 and T2. For the non-TIR subfamily, seven probes from four clades (N1 to N4) hybridized with approximately 120 positive BAC clones. These BACs (78) were analyzed by fingerprinting and assembled into at least 20 contigs for which Southern blot analysis identified 50 distinct positive bands. The combined estimate of 103 homologous RGAs is certainly an undercount of the NBS sequences in the *M. truncatula* genome. For example, as described below, some ESTs did not cosegregate with genetic markers derived from BAC clones hybridizing to probes within the same clades. In addition, the hybridization probes were not representative of clades T3, T4, and N5 to N8, be-

Table 1. Resistance gene analog-containing BACs identified by BAC library screening

Subfamily	Subgroup ^a	Probes	Positive BACs ^b
TIR ^c	T1	MtA20, MtA88, MtB82	(07G13, 24G08, 35O18), (22I12, 24K22, 44D11), (08C10, 24E08, 33I14), (12K21, 14N13, 15L04, 34F21), (38N02, 51J01, 52H11), (28B11, 36F11, 38K06, 42O05, 46D03, 76K16), (28L13, 76N03, 78B21), (58I14, 62L09), (01E19, 46J05), (19O04*, 19O17*), 11N15, 79A13, 12B16, 50O01, 17K08, 69H07, 62K19, 33K13
	T2	MtA3, MtA4, MtA6, MtA92, MtA99	(01F16, 01H02, 16A04, 11N17, 25A23, 37J05, 74B16), (20A09, 36N01, 46H17, 48N18), (01N06, 03F12, 26E21, 35A08, 78F22), (31E09, 79P21, 80D12), (18L14, 20E08, 37A07), (11B02, 23J03, 33H10, 73I08), (10P06, 18A05, 19O04*, 19O17*), 50C10
Non-TIR	N1	Mt02a, Mt08a	(18D24, 24D15, 36F01, 47H09, 80I07), (08G05, 12A16, 19F14, 32I05, 37D01, 37D11*), (01C05, 19L13, 33P15*), (13B03*, 26G03*), 01E04, 01F21*, 02B13, 03F15*, 07E07, 08M06, 14C20, 14M16, 26I20, 38L12, 40L12, 40I05*, 42C16*, 42J16*, 74O05, 80K10
	N2	Mt09a, Mt11a	(02M10, 02O18), (03N09, 05J09, 33N19), (13B03*, 26G03*), 02L03, 06M23, 08B3, 11O09, 15J09, 15J11, 38K01, 40H12, 41O18, 44M13(AQ917523), 47M11, 75D01, 75L05, 79H20, 04J02*, 35N06*, 37D11*, 42C16*, 80K22
	N3	Mt13, Mt14	(01P06, 33I23), (03N06, 11E07, 20M19, 27F23, 31E20, 36C01), (10D07, 17A18), (12K15, 23L16, 41F21), (15D08, 31E23), (21L23, 45J20), 04J02, 14D10, 14E24, 20I24, 43I21
	N4	Mt03a	(01D10, 03O07, 13L14, 21L20, 33O15, 34D20), (07H15, 35N06), (11G02, 28B12, 30A10, 31M14, 41F23, 44F10), 19D07, 01F21*, 03F15*, 40I05*, 42J16*, 33P15*, 35N06*, 45C08

^a Subgroups are assigned according to the phylogenetic analysis illustrated in Figure 1.

^b BAC clones within parentheses belong to the same contig. Underlined BACs are those that were mapped, and asterisks indicate the BACs that are positive for multiple probes.

^c TIR = Toll and Interleukin-1 receptor.

cause these clades were discovered subsequently based on BLAST analysis of the Mt GeneIndex.

More than 60% of the BAC contigs analyzed by hybridization contain multiple copies of RGAs, with as many as five copies on a single BAC clone (Fig. 3). As shown in Table 1, probes from different clades within the same subfamily (TIR or non-TIR) typically identified different sets of BACs, with occasional BACs common to distinct clades. The cross-hybridization between probes from different clades is unlikely (though still possible), considering the significant sequence divergence between clades and the high hybridization stringency used in these analyses. Instead, these results suggest the colocalization of different classes of NBS sequences within individual BACs, a conclusion that is further supported by the genetic mapping experiments described below. In contrast to the hybridization findings within either NBS subfamily, TIR and non-TIR NBS sequences were never found to be on the same BAC clones. This may indicate a lack of interspersion of TIR and non-TIR NBS sequences in the *M. truncatula* genome at the scale of individual BAC clones.

Genetic mapping of the TIR- and non-TIR NBS sequences in the *M. truncatula* genome.

With the intent of determining the distribution of individual RGAs and RGA clusters within the genome, most of the NBS-containing BAC contigs (predicted to represent 103 distinct RGAs) and 20 R-ESTs were placed on the *M. truncatula* genetic linkage map (Fig. 4). Several additional ESTs could be directly anchored to the mapped BAC contigs by virtue of their identity with BAC end sequences (Fig. 4). Together, these genetic markers represent 54 loci dispersed over seven of the eight *M. truncatula* linkage groups (LGs). On a genetic scale the majority of mapped RGAs are clustered within the genome, and in no case were genetically defined clusters found to contain both TIR and non-TIR NBS sequences, expanding on the BAC level clustering identified based on hybridization and fingerprint analysis.

The distribution of TIR-type NBS sequences was significantly skewed, with most of the mapped loci occurring on either LG 4 or LG 6. Hybridization data indicate the presence of at least eight NBS-containing sequences within the LG 4 cluster (Fig. 3). Three R-ESTs (AW696680, AW688591, and AW256656) also mapped to the same cluster. Most of the TIR type NBS sequences, however, are located on LG6, in which hybridization and mapping data identified more than 30 NBS sequences within a 15-cM segment. This extended cluster on the LG 6 cluster is apparently heterogeneous, containing diverse sequences from both the T1 and T2 clades, including five R-ESTs (AW684911, BE325283, AL369047, TC19157, and AW256570). The coincidence of R-ESTs with mapped BAC contigs indicates that many of these RGAs are transcriptionally active. By contrast, LG 6 contains relatively few mapped ESTs other than those with homology to RGAs (D. Kim and D. Cook, *unpublished data*), indicating that, while the linkage group is overrepresented for TIR RGAs, it is also significantly underrepresented for randomly selected ESTs. The LG 6 chromosome is also the shortest of the *M. truncatula* chromosomes and contains the largest proportion of heterochromatic DNA (Kulikova et al. 2001); this later feature would be consistent with an underrepresentation of transcribed genes.

In the case of non-TIR RGAs, most characterized ESTs and BAC-associated clusters mapped to LG 3, LG 5, and LG 8, although several BACs identified by probes from the N2 and N3 clades mapped as singletons in other linkage groups. Sequences from distinct clades within the non-TIR subfamily were closely linked on LG 3 and LG 5, consistent with the BAC hybridization data that indicates microheterogeneity within these clusters (Table 1). The fact that sequences are of-

ten more similar between than within the LG 5 and LG 3 clusters is consistent with these two groups of sequences having arisen by segmental duplication of the genome. In the majority of cases, R-ESTs mapped in close proximity to RGAs identified based on genomic sequence, indicating a similar genomic distribution of expressed resistance gene analogs.

DISCUSSION

Diversity of the NBS-LRR family of RGAs in legumes.

Phylogenetic analysis of hundreds of NBS sequences from several plant families reveals that most major sequence clades are family-specific (Cannon et al. 2001; Meyers et al. 1999; Pan et al. 2000b), indicating that these sequences have experienced substantial structural and functional divergence since radiation of these families. Family-specific NBS-LRR genes may also exhibit restricted taxonomic functionality, as shown for the cloned R genes *Bs2* from pepper and *RPS2* from *Arabidopsis*, which function in related species within the same family but not in species outside of the family (Tai et al. 1999). These results suggest that, within a given plant family, R genes can be considered part of a single coherent genetic system.

In the current work, we have examined the diversity and organization of RGAs within the model legume *M. truncatula* and made detailed comparisons to other species within the Papilionoid subfamily of legumes. We have determined that, although the Papilionoid legume species appear to share a similar breadth of RGA diversity, the phylogenetic relationship between individual RGAs from seven Papilionoid legume species is closely correlated with the established phylogeny of the species themselves (Doyle 1998). Thus, while the origin of the principal clades of RGAs appears to predate radiation within the Papilionoid subfamily, speciation within this group of plants is correlated with continued divergence of the RGA family. Interestingly, single genotypes of many plant pathogens are able to infect closely related legume species (P. Uribe and D. Cook, *unpublished data*), which may be an indication of taxonomically restricted R gene diversity. Conversely, the occurrence of resistance against single pathogen genotypes in closely related legume species may provide an opportunity to test the hypothesis that at least some resistance gene specificities might predate speciation events.

Well-characterized model genomes, with appropriate phylogenetic positions, may offer a means to extend the lists of RGAs in agronomically important species. In the case of legumes, representatives of most sequence types identified so far from Papilionoid crop species have also been found in *M. truncatula* (Fig. 2), establishing it as a useful model species for the NBS-LRR gene family within this subfamily of the tribe Fabaceae. The presence of RGA lineages in the *M. truncatula* genome that lack representation by sequences from other legume species could be due to insufficient sampling depth, divergence of degenerate primer target sites, or the loss or gain of particular sequence types in some taxa. In the current study, we encountered both a case in which a clade containing several *M. truncatula* sequences has no apparent soybean homologs (Fig. 2, clade j) and cases in which clades initially appearing to contain only soybean or *M. truncatula* sequences subsequently led to the recovery of *M. truncatula* or soybean homologs (Fig. 2, clades m and k, respectively). Thus, sequences from one species may provide clues to the presence of a similar set of sequences in closely related species. One might expect that species from within the same genus would exhibit an even more direct correspondence of RGA genes. We compared the TIR subfamily (Fig. 2) of RGAs between two *Medicago* species (*sativa* and *truncatula*) and determined that many branches that contain sequences from *M. truncatula* also contain *M. sativa*

sequences, with identities as high as 95%. Such sequences represent potential orthologous genes or the recent radiation from an orthologous ancestor. These results indicate not only high levels of similarity between *Medicago* RGAs but also a similar sequence diversity within the two species. Although the non-TIR subfamily has not been deeply sampled in *M. sativa*, a

similar result might be expected for this subfamily within the genus *Medicago*. A comparable situation has been reported within the Solanaceae, where potato and tomato sequences form tight clusters in a phylogenetic analysis and some tomato NBS-LRRs are more closely related to potato homologs than to other tomato-derived sequences (Pan et al. 2000b).

Fig. 4. Map positions of nucleotide binding site-containing BAC contigs and resistance gene (R) expressed sequence tags (ESTs). All markers labeled with an R- are R-gene related. R-ESTs are indicated by their GenBank accession number or tentative consensus number from the *Medicago truncatula* Gene-Index, and the mapped BAC contigs are indicated by the respective BAC IDs from which the markers are developed. Identity between ESTs and BAC end sequences is signified by an “=” sign. Names in ovals (i.e., N1 to N4 and T1 to T2) on the left side of each linkage group refer to clades. Linkage group numbers are indicated above each chromosome.

Genomic organization of NBS sequences in *M. truncatula*.

Detailed genetic and physical mapping of both TIR and non-TIR related NBS sequences in *M. truncatula* reveals that most *M. truncatula* RGAs are clustered. Some of these *M. truncatula* clusters contain sequences from only a single phylogenetic clade, while others are mixed and comprised of highly dissimilar sequences within the TIR or non-TIR subfamily. Unequal crossing over or gene conversion between related sequences is a proposed mechanism for sequence homogenization and concerted evolution within clusters; such mechanisms could have contributed to the expansion and homogeneity of clusters such as that of TIR subgroup T2 on LG 4. At the other extreme, heterogeneous RGA clusters may arise by sequence divergence due, separately or in combination, to diversifying selection after duplication, to selective gene loss, or to ectopic recombination among physically distinct clusters (Michelmore and Meyers 1998; Parniske and Jones 1999). Such mechanisms could potentially explain the structural features of the three non-TIR NBS clusters on LG 3, LG 5, and LG 8, as each of these clusters contain sequences that are cophyletic with RGAs present in at least one other cluster, while none of the clusters contain an identical set of sequence clades.

Mixed clusters of RGAs appear to be a common feature in higher plants. Nevertheless, there are dissimilarities between the structural features we have defined in *M. truncatula* and certain species examined to date. For example, in *Arabidopsis thaliana* sequences from the TIR and non-TIR groups are sometimes found together in mixed clusters (Meyers et al. 1999). Our results, which are based on a combination of genetic and physical mapping of a wide range of RGA sequences, suggest that TIR and non-TIR NBS sequences do not occur together in the *M. truncatula* genome. It seems unlikely that this type of organization is simply the result of stochastic expansion and contraction of local clusters. A similar situation may exist in soybean, as a major cluster of TIR NBS sequences maps to LG J and a non-TIR NBS cluster has been identified on LG F (Kanazin et al. 1996; Yu et al. 1996), but mixed TIR and non-TIR clusters have not been identified. Separate clustering of TIR and non-TIR NBS-LRR sequences could have contributed to the ancient divergence of these two subfamilies, for example, by restricting locally acting mechanisms for sequence homogenization such as unequal crossing over.

Syntenic relationship of RGAs between *M. truncatula* and other legumes.

R genes may have experienced more rapid structural and functional divergence and genomic reorganization than the rest of the genome (Leister et al. 1998), a feature that might be expected to complicate the identification of syntenic clusters of R genes between genomes. For example, predicting gene lineages and distinguishing orthologs from paralogs might be especially difficult for large and rapidly evolving multigene families such as R genes (Michelmore 2000); similarly, the presence of null alleles for some R genes would further complicate this task (Grant et al. 1998; Stahl et al. 1999). It is, therefore, not surprising that the chromosomal positions of RGAs seem not to be conserved between grass species, although remarkable conservation of gene order has been observed for other genes within these related genomes (Leister et al. 1998). This finding suggests that the creation and maintenance of R gene organization may be partly independent of organization of the surrounding genome. In contrast to the results obtained for the Poaceae, conserved syntenic positions of R genes have been observed within the Solanaceae family (Grube et al. 2000; Pan et al. 2000b). The contrasting results obtained for members of the Poaceae versus the Solanaceae tribes may be due to the different evolutionary histories of these plant families (Pan et al.

2000b). Even in cases where the position of clusters is conserved after speciation, however, the fine structure of these clusters and the functions of the genes they contain may not be maintained (Grube et al. 2000).

M. truncatula is expected to provide a point of comparison for the structure of related legume genomes. Although limited information is available so far, there is strong evidence for a conserved genomic location of several RGA loci between *M. truncatula* and soybean and between *M. truncatula* and pea. In each of these cases, not only is synteny evident based on reference to adjacent low copy genes, but also the RGAs themselves have a conserved phylogenetic position. For example, soybean LG F contains clusters of various disease resistance loci, as well as numerous non-TIR RGAs (Ashfield et al. 1998; Yu et al. 1996). Restriction fragment length polymorphism (RFLP) marker B212 is closely linked to this R gene-rich region in soybean. Interestingly, *M. truncatula* BAC clone 03N06, which was identified by hybridization with soybean non-TIR NBS sequences, is closely linked to a homolog of soybean B212 (i.e., marker DK505) on *M. truncatula* LG 8 (Fig. 4). In both species, the syntenic RGAs are members of the non-TIR clade m (Fig. 2). This result is suggestive of a conserved location of these non-TIR RGAs in these two legume genomes. Similarly, soybean LG J contains a cluster of several R gene loci, as well as one nodulation gene (*Rj2*) and many TIR-related NBS sequences (Kanazin et al. 1996; Polzin et al. 1994). This cluster is closely linked to RFLP marker A233. The homologous marker of A233 in *M. truncatula* (DK321) (D. Kim and D. Cook, unpublished data) is also linked to a mixed TIR cluster in *M. truncatula* LG 6 (Fig. 4). The corresponding RGAs from soybean and *M. truncatula* are members of TIR clade a, consistent with a syntenic relationship between the two clusters. The LG 6 TIR cluster of *M. truncatula* represents an extended cluster of at least 30 RGAs that spans the majority of LG 6. This organization is reminiscent of the *Dm3* cluster characterized in lettuce, where at least 24 resistance gene homologs are spread over at least 3.5 Mb (Meyers et al. 1998).

Recently, *M. truncatula* LG 5 has been shown to have a high degree of macrosynteny and microsynteny with pea LG 1 (Geurts et al. 2000). LG 1 of pea is populated with many symbiosis- and resistance-related genes, in particular a non-TIR RGA cluster (AW123698) with very high homology to *M. truncatula* Mt03 type non-TIR NBS sequences (Timmerman-Vaughan et al. 2000). Interestingly, a cluster of Mt03-type non-TIR NBS sequences is also located on LG 5 of *M. truncatula*, adjacent to cross-species reference marker ENOD40 (Fig. 4), indicating that these clusters of phylogenetically related non-TIR RGAs are also syntenic between the two genomes.

Similar syntenic relationships seem even more likely to be identified for the genome of the closely related forage legume alfalfa (*M. sativa*). *M. truncatula* and diploid *M. sativa* have the same number of chromosomes ($2n = 16$), and the global syntenic relationships between the two genomes have been well established (G. Kiss and D. Cook, unpublished data). The frequent presence of paired *Medicago* sp. RGA sequences in the phylogenetic analysis is entirely consistent with this expectation and work is currently underway to test this prediction.

A common theme for several of the *M. truncatula* genome projects is a strong focus on plant-microbe interactions (Cook 1999; Frugoli and Harris 2001). In fact, many of the *M. truncatula* EST libraries have been developed from tissues challenged with pathogenic or symbiotic microorganisms. Moreover, there is a large increase in the numbers of researchers focusing on other pathogen and insect pests of *M. truncatula*, with the specific intent of identifying resistance phenotypes. As stated above, many of the pathogens of *M. truncatula* are also patho-

gens of closely related legumes. For example, in an ongoing study of genetic resources in *M. truncatula*, functional resistance was identified against *Erysiphe pisi*, the powdery mildew pathogen of *Pisum sativum* (pea). The phenotype mapped to the LG 5 non-TIR cluster in *M. truncatula* (N. D. Young, *personal communication*), which shares both synteny and close phylogenetic relatedness to a resistance gene cluster in pea. This observation suggests two important opportunities. First, it should be possible to clone R genes that are active against pathogens of crop legume species in *M. truncatula*. Second, due to the close phylogenetic relationship of RGA sequences between these species (Fig. 2), it is likely that functional R genes can be moved across species boundaries by transgenic approaches.

The results of this current study add a potentially important component to knowledge of plant-microbe interactions in *M. truncatula*, namely a well-defined genomic structure and phylogenetic context for resistance gene analogs. Taken together, these resources should greatly facilitate the study of this important aspect of legume biology in *M. truncatula*.

MATERIALS AND METHODS

PCR amplification, database searching, and sequence analysis.

Two pairs of degenerate primers were used for PCR amplification of NBS sequences from the *M. truncatula* genome. The first pair of primers, F1 and R1, was designed based on conserved motifs in the NBS region of both the TIR and non-TIR NBS-LRR family of R genes. The forward primer F1, 5'-GGI-GGIRTIGGIAAIAAC-3', corresponds to the P-loop motif (GGVGKTT) in the sense direction, while the reverse primer R1, 5'-IARIGYIARGGIARICC-3', corresponds to the GLPL motif in the antisense direction. The second pair of primers, F2 and R2, was designed for amplification of NBS sequences that occur in R genes lacking N-terminal TIR domains. The forward primer F2 (sense), 5'-GGIGGIGTIGGIAAIAAC-3', is similar to the F1 primer, whereas the reverse primer R2 (antisense), 5'-GGRAAIARISHRCARTAVIRAARC-3', was designed based on the CFLYCALFP motif that is well conserved among non-TIR but missing in TIR NBS sequences (Meyers et al. 1999).

Genomic DNA from *M. truncatula* genotype Jemalong A17 was used as template for PCR amplification. PCR products were cloned into the pPCR-Script Amp SK (+) vector (Stratagene, La Jolla, CA, U.S.A.) or P gem-T Easy vector (Promega, Madison, WI, U.S.A.) and transformed into competent cells of DH10B (Gibco/BRL, Rockville, MD, U.S.A.) according to the manufacturer's protocols. Plasmid DNA was extracted using the QIAprep Spin Miniprep Kit (Qiagen, Valencia, CA, U.S.A.) and sequenced with an ABI377 automated sequencer and the ABI Prism BigDye terminator cycle sequencing ready reaction kit (Applied Biosystems, Foster City, CA, U.S.A.) following the recommended protocols. DNA inserts were sequenced using primers complementary to the polylinker T3, T7, or M13 sites.

Two additional *M. truncatula* sequences in the non-TIR subfamily were obtained based on hybridization to soybean NBS sequences (GenBank accession numbers AF301766, AZ301628, and AZ301636) of the non-TIR subfamily (S. Penuela and N. D. Young, *unpublished data*). The first, called "Mt13", was subcloned from an *M. truncatula* BAC clone 27F23. The second sequence, "Mt14", was amplified and cloned from a BAC clone (10D07) using the F2 and R2 primers. All RGA sequences reported in this paper were submitted to GenBank with accession numbers AZ773476 to AZ773483 and AZ773496 to AZ773519.

We also identified NBS sequences from the TIGR *Medicago truncatula* GeneIndex (MtGI) database, using the TBLASTN

algorithm (Altschul et al. 1997). Fourteen known NBS-encoding R gene protein sequences were used to query the database: N (A54810), L6 (U27081), M (U73916), RPP5 (AAF08790), RPS4 (AJ243468), RPP1 (AF098962), I2 (AF118127), Mi (AF039682), RPM1 (A57072), RPP8 (AAC83165), RPS2 (A54809), RPS5 (AAC26126), Xa1 (AB002266), and Pi-b (AB013448). To avoid identification of sequences with similarity to only the LRR regions of R genes, the LRR region of the query sequences was removed. The searches were performed on July 23, 2001, when the MtGI contained 127,624 submissions.

DNA sequences were compiled and translated using Sequencher 3.1.1 (Gene Codes, Ann Arbor, MI, U.S.A.). Sequence alignments and phylogenetic analyses were performed on amino acid translations of the DNA sequence data using the ClustalX algorithm (Thompson et al. 1994, 1997). Phylogenetic trees were constructed using the neighbor-joining method as implemented in ClustalX with 1,000 bootstrap sampling steps (Saitou and Nei 1987). Substitution rates were corrected for multiple hits using Kimura's protein distance correction.

The other legume sequences used for comparative phylogenetic analysis were from *M. sativa*, *Pisum sativum*, *G. max*, *Phaseolus vulgaris*, *Cicer arietinum*, and *Cajanus cajan*. All sequences were either PCR-derived or ESTs retrieved from the NCBI nonredundant database. Sequences were chosen to represent the range of distinct sequence classes that have been identified within the given species.

BAC library screening, fingerprinting, and contig assembly.

High-density filters containing a 5× *M. truncatula* BAC library of 30,720 clones (Nam et al. 1999) were obtained from the Clemson University Genome Center. These filters were screened by hybridization with NBS clones selected to represent the primary homology groups as determined by phylogenetic analyses, described above. BAC fingerprinting and contig assembly were according to Marra and associates (1997). Prior to further analysis, the assignment of individual BAC clones to particular NBS homology groups was checked by Southern blot analysis. For this purpose, DNA fingerprinting gels were blotted to nylon membranes and hybridized with a panel of NBS clones following standard protocols.

BAC end sequencing, marker development, and genetic mapping.

Typically several BAC ends within a BAC contig were sequenced to provide a basis for genetic marker development. The BAC end sequencing primers were designated SQ-BAC-L (5'-AACGCCAGGGTTTCCAGTCACGACG-3') and SQ-BAC-R (5'-ACACAGGAAACAGCTATGACCATGATTACG-3'), which correspond to the left and right flanking ends of the BAC vector, respectively. Sequencing was performed on an ABI 377 (Applied Biosystems, Foster City, CA, U.S.A.) using the ABI Prism BigDye terminator cycle sequencing ready reaction kit (Applied Biosystems).

High-quality BAC end sequences were used as templates for primer design using the Lasergene PrimerSelect software package (DNASar, Madison, WI, U.S.A.). Individual primer pairs were then used for PCR amplification of genomic DNA from *M. truncatula* genotypes A17 and A20, two parents of an F2 mapping population (Penmetsa and Cook 2000). Each pair of corresponding sequences from genotypes A17 and A20 were aligned using Sequencher software (Gene Codes) to detect single nucleotide polymorphisms (SNPs). Once putative SNPs were detected, the sequences were transferred to DNA Strider 1.2 for restriction site identification. Restriction sites coincident with SNPs were used to develop cleavable amplified polymorphic sequence (CAPS) markers for genetic mapping. A

similar strategy was used for developing CAPS markers from known *M. truncatula* EST sequences. The R-ESTs chosen for mapping were either complete NBS sequences or partial NBS motifs adjacent to other domains (e.g., TIR or CC) that are diagnostic for the TIR or non-TIR type sequences.

The *M. truncatula* mapping population consisted of 93 F2 individuals from a cross of genotypes A17 × A20 (Penmetza and Cook 2000). DNA from the two parents and an F1 individual served as controls for the expected genotypes. PCR reactions were performed in a total volume of 10 µl containing 20 ng of genomic DNA, 1× PCR reaction buffer, 2.5 mM MgCl₂, 0.25 mM of each dNTP, 2.5 pmole of each primer, and 0.5 units of HotStartTaq DNA polymerase (Qiagen). The temperature program consisted of 15 min at 95°C, followed by 35 cycles of 94°C for 30 s, 55 to 60°C (dependent on primer T_m) for 30 s, and 72°C for 30 s with a single final extension of 7 min at 72°C. Following PCR, 1.5 µl of amplification product was transferred to a fresh 96-well plate for digestion with restriction enzyme. Digestion products were resolved on agarose gels of appropriate percentage and scored for the respective homozygous parental and heterozygous genotypes. Linkage mapping was performed using MAPMAKER Version II for Macintosh.

ACKNOWLEDGMENTS

This research was supported by grant no. 9872664 from the National Science Foundation Plant Genome Research Program to DRC and NDY and by a United States Department of Agriculture training grant to SBC. The authors wish to thank D. Kim, S. Penuela, and G. Kiss for technical assistance and advice.

LITERATURE CITED

Aarts, N., Metz, M., Holub, E., Staskawicz, B. J., Daniels, M. J., and Parker, J. E. 1998. Different requirements for *EDS1* and *NDR1* by disease resistance genes define at least two R gene-mediated signaling pathways in *Arabidopsis*. *Proc. Natl. Acad. Sci. U.S.A.* 95:10306-10311.

Altschul, S. F., Madden, T. L., Schaffer, A. A., Zhang, J., Zhang, Z., Miller, W., and Lipman, D. J. 1997. Gapped BLAST and PSI-BLAST: A new generation of protein database search programs. *Nucleic Acids Res.* 25:3389-3402.

The Arabidopsis Genome Initiative. 2000. Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. *Nature* 408:796-815.

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

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

Bent, A. F. 1996. Plant disease resistance genes: Function meets structure. *Plant Cell* 8:1757-1771.

Cannon, S. B., Zhu, H., Baumgarten, A. M., Spangler, R., May, G., Cook, D. R., and Young, N. D. Diversity, distribution, and ancient taxonomic relationships within the TIR and non-TIR NBS-LRR resistance gene subfamilies. *J. Mol. Evol.*, in press.

Century, K. S., Shapiro, A. D., Repetti, P. P., Dahlbeck, D., Holub, E., and Staskawicz, B. J. 1997. *NDR1*, a pathogen-induced component required for *Arabidopsis* disease resistance. *Science* 278:1963-1965.

Cook, D. R. 1999. *Medicago truncatula*-A model in the making! *Curr. Opin. Plant Biol.* 2:301-304.

Dangl, J., and Holub, E. 1997. La dolce vita: A molecular feast in plant-pathogen interactions. *Cell* 91:17-24.

Doyle, J. J. 1998. Phylogenetic perspectives on nodulation: Evolving views of plants and symbiotic bacteria. *Trends Plant. Sci.* 3:473-478.

Ellis, J., Dodds, P., and Pryor, T. 2000. Structure, function and evolution of plant disease resistance genes. *Curr. Opin. Plant Biol.* 3:278-284.

Frugoli, J., and Harris, J. 2001. *Medicago truncatula* on the move! *Plant Cell* 13:458-463.

Geurts, R., Gualtieri, G., Kulikova, O., Kim, D. J., Cook, D. R., and

Bisseling, T. 2000. A region on the upper arm of chromosome 5 of *Medicago truncatula* is highly syntenic to the *sym2* region of pea. Pages 309-310 in: Nitrogen Fixation: From Molecules to Crop Production. Kluwer Academic Publishers, Dordrecht, The Netherlands.

Grant, M. R., McDowell, J. M., Sharpe, A. G., de Torres Zabala, M., Lydiate, D. J., and Dangl, J. L. 1998. Independent deletions of a pathogen-resistance gene in *Brassica* and *Arabidopsis*. *Proc. Natl. Acad. Sci. U.S.A.* 95:15843-15848.

Grube, R. C., Radwanski, E. R., and Jahn, M. 2000. Comparative genetics of disease resistance within the solanaceae. *Genetics* 155:873-887.

Hammond-Kosack, K. E., and Jones, J. D. G. 1997. Plant disease resistance genes. *Annu. Rev. Plant Physiol. Plant Mol. Biol.* 48:575-607.

Kanazin, V., Marek, L. F., and Shoemaker, R. C. 1996. Resistance gene analogs are conserved and clustered in soybean. *Proc. Natl. Acad. Sci. U.S.A.* 93:11746-11750.

Kulikova, O., Gualtieri, G., Geurts, R., Kim, D. J., Cook, D., Huguet, T., de Jong, J. H., Fransz, P. F., and Bisseling, T. 2001. Integration of the FISH pachytene and genetic maps of *Medicago truncatula*. *Plant J.* 27:49-58.

Leister, D., Ballvora, A., Salamini, F., and Gebhardt, C. 1996. A PCR-based approach for isolating pathogen resistance genes from potato with potential for wide application in plants. *Nat. Genet.* 14:421-429.

Leister, D., Kurth, J., Laurie, D. A., Yano, M., Sasaki, T., Devos, K., Graner, A., and Schulze-Lefert, P. 1998. Rapid reorganization of resistance gene homologues in cereal genomes. *Proc. Natl. Acad. Sci. U.S.A.* 95:370-375.

Lynch, M., and Conery, J. S. 2000. The evolutionary fate and consequences of duplicate genes. *Science* 290:1151-1155.

Marra, M. A., Kucaba, T. A., Dietrich, N. L., Green, E. D., Brownstein, B., Wilson, R. K., McDonald, K. M., Hillier, L. W., McPherson, J. D., and Waterston, R. H. 1997. High throughput fingerprint analysis of large-insert clones. *Genome Res.* 7:1072-1084.

Meyers, B. C., Chin, D. B., Shen, K. A., Sivaramakrishnan, S., Lavelle, D. O., Zhang, Z., and Michelmore, R. W. 1998. The major resistance gene cluster in lettuce is highly duplicated and spans several megabases. *Plant Cell* 10:1817-1832.

Meyers, B. C., Dickerman, A. W., Michelmore, R. W., Sivaramakrishnan, S., Sobral, B. W., and Young, N. D. 1999. Plant disease resistance genes encode members of an ancient and diverse protein family within the nucleotide-binding superfamily. *Plant J.* 20:317-332.

Michelmore, R. 2000. Genomic approaches to plant disease resistance. *Curr. Opin. Plant Biol.* 3:125-131.

Michelmore, R. W., and Meyers, B. C. 1998. Clusters of resistance genes in plants evolve by divergent selection and a birth-and-death process. *Genome Res.* 8:1113-1130.

Nam, Y., Penmetza, R. V., Endre, G., Uribe, P., Kim, D., and Cook, D. R. 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.

Pan, Q., Wendel, J., and Fluhr, R. 2000a. Divergent evolution of plant NBS-LRR resistance gene homologues in dicot and cereal genomes. *J. Mol. Evol.* 50:203-213.

Pan, Q., Liu, Y. S., Budai-Hadrian, O., Sela, M., Carmel-Goren, L., Zamir, D., and Fluhr, R. 2000b. Comparative genetics of nucleotide binding site-leucine rich repeat resistance gene homologues in the genomes of two dicotyledons: Tomato and *Arabidopsis*. *Genetics* 155:309-322.

Parker, J. E., Holub, E. B., Frost, L. N., Falk, A., Gunn, N. D., and Daniels, M. J. 1996. Characterization of *eds1*, a mutation in *Arabidopsis* suppressing resistance to *Peronospora parasitica* specified by several different RPP genes. *Plant Cell* 8:2033-2046.

Parniske, M., and Jones, J. D. G. 1999. Recombination between diverged clusters of the tomato *cf-9* plant disease resistance gene family. *Proc. Natl. Acad. Sci. U.S.A.* 96:5850-5855.

Penmetza, R. V., and Cook, D. R. 2000. Production and characterization of diverse developmental mutants of *Medicago truncatula*. *Plant Physiol.* 123:1387-1397.

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

Richter, T. E., and Ronald, P. C. 2000. The evolution of disease resistance genes. *Plant Mol. Biol.* 42:195-204.

Rivkin, M. I., Vallejos, C. E., and McClean, P. E. 1999. Disease-resistance related sequences in common bean. *Genome* 42:41-47.

Saitou, N., and Nei, M. 1987. The neighbor-joining method: A new method for reconstructing phylogenetic trees. *Mol. Biol. Evol.* 4:406-425.

- Stahl, E. A., Dwyer, G., Mauricio, R., Kreitman, M., and Bergelson, J. 1999. Dynamics of disease resistance polymorphism at the *Rpm1* locus of *Arabidopsis*. *Nature* 400:667-671.
- Tai, T. H., Dahlbeck, D., Clark, E. T., Gajiwala, P., Pasion, R., Whalen, M. C., Stall, R. E., and Staskawicz, B. J. 1999. Expression of the *Bs2* pepper gene confers resistance to bacterial spot disease in tomato. *Proc. Natl. Acad. Sci. U.S.A.* 96:14153-14158.
- Thompson, J. D., Higgins, D. G., and Gibson, T. J. 1994. CLUSTAL W: Improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res.* 22:4673-4680.
- Thompson, J. D., Gibson, T. J., Plewniak, F., Jeanmougin, F., and Higgins, D. G. 1997. The CLUSTAL_X windows interface: Flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Res.* 25:4876-4782.
- Timmerman-Vaughan, G. M., Frew, T. J., and Weeden, N. F. 2000. Characterization and linkage mapping of R-gene analogous DNA sequences in pea (*Pisum sativum* L.). *Theor. Appl. Genet.* 101:241-247.
- van der Vossen, E. A., van der Voort, J. N., Kanyuka, K., Bendahmane, A., Sandbrink, H., Baulcombe, D. C., Bakker, J., Stiekema, W. J., and Klein-Lankhorst, R. M. 2000. Homologues of a single resistance-gene cluster in potato confer resistance to distinct pathogens: A virus and a nematode. *Plant J.* 23:567-576.
- Wei, F., Gobelman-Werner, K., Morroll, S. M., Kurth, J., Mao, L., Wing, R., Leister, D., Schulze-Lefert, P., and Wise, R. P. 1999. The *Mla* (powdery mildew) resistance cluster is associated with three NBS-LRR gene families and suppressed recombination within a 240-kb DNA interval on chromosome 5S (1HS) of barley. *Genetics* 153:1929-1948.
- Young, N. D. 2000. The genetic architecture of resistance. *Curr. Opin. Plant. Biol.* 3:285-290.
- Yu, Y. G., Buss, G. R., and Maroof, M. A. 1996. Isolation of a super-family of candidate disease-resistance genes in soybean based on a conserved nucleotide-binding site. *Proc. Natl. Acad. Sci. U.S.A.* 93:11751-11756.

AUTHOR-RECOMMENDED INTERNET RESOURCES

The Institute for Genomic Research *Medicago truncatula* Gene Index:
<http://www.tigr.org/tdb/mtgi>