Available online at www.sciencedirect.com
SCIENCE DIRECT®
Molecular Phylogenetics and Evolution 31 (2004) 153–163
MOLECULAR PHYLOGENETICS AND EVOLUTION
www.elsevier.com/locate/ympev

Comparative analysis of NBS domain sequences of NBS-LRR disease resistance genes from sunflower, lettuce, and chicory

Alex Plocik, Jenn Layden, and Rick Kesseli*

Department of Biology, University of Massachusetts, Boston, MA 02125-3393, USA

Received 25 March 2003; revised 2 July 2003

Abstract

Plant resistance to many types of pathogens and pests can be achieved by the presence of disease resistance (R) genes. The nucleotide binding site-leucine rich repeat (NBS-LRR) class of R-genes is the most commonly isolated class of R-genes and makes up a super-family, which is often arranged in the genome as large multi-gene clusters. The NBS domain of these genes can be targeted by polymerase chain reaction (PCR) amplification using degenerate primers. Previous studies have used PCR derived NBS sequences to investigate both ancient R-gene evolution and recent evolution within specific plant families. However, comparative studies with the Asteraceae family have largely been ignored. In this study, we address recent evolution of NBS sequences within the Asteraceae and extend the comparison to the *Arabidopsis thaliana* genome. Using multiple sets of primers, NBS fragments were amplified from genomic DNA of three species from the family Asteraceae: *Helianthus annuus* (sunflower), *Lactuca sativa* (lettuce), and *Cichorium intybus* (chicory). Analysis suggests that Asteraceae species share distinct families of R-genes, composed of genes related to both coiled-coil (CC) and toll-interleukin-receptor homology (TIR) domain containing NBS-LRR R-genes. Between the most closely related species, (lettuce and chicory) a striking similarity of CC subfamily composition was identified, while sunflower showed less similarity in structure. These sequences were also compared to the *A. thaliana* genome. Asteraceae NBS gene subfamilies appear to be distinct from *Arabidopsis* gene clades. These data suggest that NBS families in the Asteraceae family are ancient, but also that gene duplication and gene loss events occur and change the composition of these gene subfamilies over time.

Keywords: Disease resistance genes; Molecular evolution; Asteraceae; NBS-LRR; Sunflower; Lettuce; Chicory

1. Introduction

Plants defend themselves against pathogen attacks with disease resistance (R) genes. The protein products of R-genes are believed to encode receptors that specifically recognize pathogen avirulence (Avr) proteins or guard critical components of the plant's physiological processes (Dangl and Jones, 2001). An incompatible interaction of host R-gene protein products with pathogen Avr proteins produces a defense response termed the hypersensitive response. This response impedes pathogen progression via a variety of mechanisms, including localized programmed cell death which may involve signaling components analogous to animal apoptosis (Lam et al., 2001). In a compatible interaction, pathogen growth proceeds and infection occurs.

The protein structures of disease resistances genes are diverse. Recent cloning of R-genes from a variety of plants has identified four structural classes of proteins containing combinations of domains (reviewed in Hammond-Kosack and Jones, 1997). Most R-genes contain leucine-rich repeat (LRR) domains attached to either nucleotide binding site (NBS) or transmembrane domains. Two classes of R-genes with transmembrane and extracellular LRR domains exist, one composed only of these two domains, and another which also contains an intracellular protein kinase domain. These classes share structural similarity with proteins also involved in development (Ellis et al., 2000).

The NBS-LRR resistance genes appear to code for intracellular receptors that are composed of a variable N-terminal domain followed by the NBS and LRR

* Corresponding author. Fax: 1-617-287-6650.
E-mail address: rick.kesseli@umb.edu (R. Kesseli).

1055-7903/$ - see front matter © 2003 Elsevier Inc. All rights reserved.
doi:10.1016/S1055-7903(03)00274-4

domains. The variable N-terminal region often includes either a coiled-coil (CC) or toll-interleukin receptor (TIR) domain. The LRR domain may be involved in protein–protein interactions, conferring recognition of pathogen Avr proteins (Dixon et al., 1996; Meyers et al., 1998). The NBS is the signaling domain, which forms an ATP nucleotide binding pocket (Tameling et al., 2002). It is highly conserved, containing several motifs that are strictly ordered (Meyers et al., 1999). Degenerate PCR using primers targeted to these conserved motifs is a rapid and facile strategy for the isolation of NBS sequences linked to disease resistance genes (Kanazin et al., 1996; Leister et al., 1996; Shen et al., 1998; Yu et al., 1996).

R-gene evolution has been proposed to involve several molecular mechanisms including unequal crossing-over and gene conversion. Since R-genes encode a diverse array of proteins maintaining specific resistances to different pathogens, a birth-and-death pattern has probably been more common than concerted evolution (Michelmore and Meyers, 1998). As with animal immune systems (Nei et al., 1997), gene duplication, point mutation, and gene loss may be the primary mechanisms by which R-genes evolve. Following gene duplications, diversifying selection on the LRR domain may create novel resistance specificities and be part of the host's arsenal in an arms race with pathogens (reviewed in Michelmore and Meyers, 1998).

The Asteraceae family is a highly successful lineage of plants that includes several agriculturally important species, including artichoke, sunflower, lettuce, chicory, and endive. Within the family, sunflower (*Helianthus annuus*) belongs to one subfamily, while lettuce (*Lactuca sativa*) and chicory (*Cichorium intybus*) are closely related and belong to one tribe of a second subfamily. Previous studies have investigated groups of NBS-LRR R-genes isolated by PCR from lettuce (Shen et al., 1998) and from sunflower (Gedil et al., 2001). Also, the largest cluster of disease resistance genes known in plants has been carefully characterized in lettuce (Meyers et al., 1998). In this study, we introduce NBS data from a third species, chicory. We report the identification of distinct NBS-LRR gene subfamilies with closely related homologs common to the three species of the Asteraceae. Also, a comparison of R-genes from *Arabidopsis thaliana* with those from the Asteraceae indicates that some gene families are old with homologs present in the common ancestor of dicots while others are young with homologs only in the progenitor of chicory and lettuce. The identification of potential gene loss and duplication events along with their role in R-gene evolution is discussed.

2. Materials and methods

2.1. Genomic DNA templates

Seed for *C. intybus* cv. Di Castelfranco and *L. sativa* cv. Salinas were provided by Ed Ryder (USDA Salinas, CA) and Richard Michelmore (University of California, Davis, CA), respectively. Genomic DNA, extracted by a modified CTAB method (Bernatzky and Tanksley, 1986), was used to amplify chicory and lettuce NBS sequences. *H. annuus* cv. RHA280 genomic DNA, provided by Dr. Steve Knapp (Oregon State University, Corvallis), was used to amplify sunflower NBS sequences.

2.2. Primer design

Primers were designed to amplify sequences similar to previously isolated NBS subfamilies in sunflower and lettuce. Two groups of primer sets were chosen (Table 1). Primers were designed to use the same forward primer, which is not subfamily-specific and targets a highly conserved motif (p-loop), combined with different reverse primers that target less conserved motifs. First, degenerate primer sets, F4RGC (Leister et al., 1996) and R4RGC1 (Yu et al., 1996), which were previously used to amplify NBS sequences in sunflower (Gedil et al., 2001), were used to amplify sequences in chicory and lettuce. Second, Asteraceae specific degenerate primers R4RGC2, R4RGC3A, and R4RGC3B were designed from previously reported NBS families isolated from lettuce (Shen et al., 1998) to amplify sequences from chicory and sunflower.

Attempting to preferentially amplify sequences related to RGC2 paralogs from lettuce, we designed the R4RGC2 primer from a protein alignment of RGC2 sequences obtained from the National Center for Biotechnology Information (NCBI) database as previously reported (Meyers et al., 1998). The same approach was not applicable for RGC3 since closely related paralogs

Table 1
Primer sequences used for NBS fragment amplification

Primer	Sequence[a]	Size (bp)	Motif targeted	Deg.
F4RGC	GGIGGIGTIGGIAAIACIAC	20	GGVGKTT	4096
R4RGC1	YCTAGTTGTRAYDATDAYYYTRC	23	S(K,R)(I,V)I(I,V)TTR	1152
R4RGC2	CATNGTYTTDATNGCDATNGG	21	PIAIKTM	1152
R4RGC3A	CCRCAYTCYTCNACDATRTC	20	DIVEECG	192
R4RGC3B	GGCATYTCNCKDATYTTRAA	20	FKIREMP	192

[a] Codes for degenerate positions are I, Inosine; R, A/G; Y, C/T; K, G/T; N, A/G/C/T; D, A/G/T; W, A/C; and S, G/C.

are not known, presumably because it is of low copy number (Shen et al., 1998). Therefore, two primers, R4RGC3A and R4RGC3B, were designed for a low degeneracy region upstream to the GLPL motif.

2.3. Amplification and subcloning of genomic NBS fragments

Genomic DNA (25 ng) was amplified using the forward and reverse primers. The forward primer, F4RGC1, targeted the GGVGKTT p-loop motif and was used in combination with a variety of reverse primers (Table 1) to amplify preferentially members of specific NBS-LRR subfamilies. Lettuce amplifications using primer sets R4RGC2, 3A, and 3B were performed as positive controls that were expected to produce sequences similar or identical to sequences from which they were designed.

PCR mixtures contained 25 ng of template DNA, 1 U *Taq* polymerase, 2.0 mM $MgCl_2$, 0.2 mM dNTPs, and 0.4 μM of each primer. Cycling conditions were as follows: initial denaturation at 95 °C for 3 min, 35 cycles of 94 °C for 30 s, 50 °C for 30 s, 72 °C for 45 s, and a final extension of 72 °C for 2 min. PCR products were size separated on a 1.0% agarose gel and the appropriately sized band was extracted (Qiagen). Purified fragments were subcloned into pGEM t-easy (Promega) and transformed into XL1-Blue MRF′ cells (Stratagene) for blue/white colony selection. Randomly picked colonies with inserts were grown and plasmid mini-prepped (Qiagen) for sequence analysis.

2.4. Sequencing

Sequencing reactions were prepared by cycle sequencing from the M13 forward or reverse primer with DyeDeoxy Terminator Cycle Sequencing Kit v3.0 containing Amplitaq FS DNA polymerase and analyzed by an ABI 377 automated sequencer (Perkin–Elmer Applied Biosystems, Foster City, CA, USA). Sequences were manually edited and cleaned of vector and primer sequences using Sequencher (Genecodes, Ann Arbor, MI). Identical clones were recovered, thus error due to *Taq* was low and estimated to be less than 0.1%. To avoid problems in the analyses caused by *Taq* error, sequences greater than 99% identical were considered to be the same and the consensus sequence from multiple clones was used.

2.5. Similarity searches for NBS homologues to isolated sequences

BLAST version 2.1.3 (Altschul et al., 1997) was used to search the Genbank non-redundant nucleotide and protein databases and the TAIR AGI total protein database. Protein similarity was determined using tBLASTx and with BLASTx without the low complexity filter and requiring an expected value of 0.0001. All other values were set to default. Other resistance gene candidates (RGCs) used in the analysis were retrieved from Genbank and identified by protein searches using isolated RGCs as queries.

2.6. Sequence alignment and phylogenetic analyses

Since NBS domains have a strict backbone of ordered motifs, consistent alignments of conceptual translations with minimal manual corrections were possible. Phylogenetic reconstructions were performed by MEGA version 2.1 (Kumar et al., 2001) using CLUSTALX 1.8 (Thompson et al., 1997) derived alignments to produce Neighbor-joining trees (Saitou and Nei, 1987) with 1000 bootstraps. Branch lengths were assigned by pair-wise calculations of genetic distances and missing data were treated by pair-wise deletions of gaps. Data were incorporated from either the p-loop to RNBS-B motif (~100 amino acids) or p-loop to GLPL motif (~160 amino acids) where appropriate; both were analyzed when possible. Trees were outgroup rooted with the NB-ARC domain of human apoptosis activating factor-1 (Accession # AAD34016), which contains homologous motifs with the NBS domain NBS-LRR disease resistance genes (van der Biezen and Jones, 1998).

A phylogenetic analysis of closely related lettuce NBS sequences, corresponding to the *Dm3* cluster of lettuce (Meyers et al., 1998), was performed incorporating closely related chicory sequences outgroup rooted by a related clade of chicory and sunflower sequences and of lettuce sequences not reported to be linked to *Dm3*. Nucleotides were aligned by an amino acid CLUSTALX alignment of conceptual translations. A neighbor-joining tree calculated by Kimura's 2-parameter model (Kimura, 1983) was applied incorporating all codon sites. Missing data were treated by pair-wise deletion of gaps.

2.7. Nomenclature

The first two letters (HA, LS, or CI) of the sequences discussed in this study indicate the genus and species. Notation following the name indicates the source of the sequence, NCBI or previously unpublished from the Kesseli lab (RVK).

3. Results

3.1. Amplification, identification, and characterization of NBS fragments

With degenerate primers (Table 1), fragments of predicted sizes were amplified, extracted and subcloned

for sequencing from lettuce, chicory and sunflower (Table 2). Care was taken specifically to analyze sequences generated by comparable methods. Under optimized conditions, the most prominent band amplified for each of the three species was always the expected size, however F4RGC apparently functioned as both the forward and reverse primer in lettuce and amplified an additional 1.3 Kb fragment. Sequencing analysis showed this band to be a large NBS fragment containing a small intron. Despite this product in lettuce, predicted sizes were also present and extractable.

The majority, 53 of 57, of unique clones isolated from bands of predicted sizes appeared to encode the NBS domains of NBS-LRR disease resistance genes. Four sequences encoded open reading frames, but were not significantly similar to the NB-ARC domain and did not contain any recognizable NBS motifs. These sequences either encoded proteins related to transposable elements or lacked significant similarity to proteins in the NCBI NR database. These amplified reading frames were often encoded in the reverse frame relative to the forward primer, indicating that these were misprimed artifacts and not generated by NBS related motifs. Misprimed sequences were not included in phylogenetic analyses.

Among the NBS related sequences, one chicory and one sunflower sequence contained frame shift mutations, while another sunflower sequence contained a stop codon (Table 2). These results are consistent with previous amplifications of NBS sequences from sunflower (Gedil et al., 2001). Putative pseudogene sequences were not included further in the analysis although they were clearly related to NBS sequences.

The 50 remaining unique NBS sequences (21 chicory, 20 lettuce, and 9 sunflower; Table 2) were identified as resistance gene candidates (RGCs) by the presence of an uninterrupted open reading frame and by characteristic R-gene NBS motifs (Meyers et al., 1999; Pan et al., 2000a). Further, conceptual translations were used to query the NCBI non-redundant, conserved domain, and the *Arabidopsis* Genome Initiative protein databases. All suspected NBS sequences retrieved matches to NBS sequences from *Arabidopsis* and other plant species with highly significant *e*-values (<0.0001).

As expected, the R4RGC1 primer, which was not designed to distinguish CC and TIR classes of genes,

Table 2
Results of PCR amplifications

	Reverse primer[a]	~Band size (bp)	NBS sequences	Pseudogenes		Classes amplified
			Unique	Stop	Frame	
Chicory	R4RGC1	330	12[b]		1	TIR/CC
	R4RGC2	550	6			CC
	R4RGC3A	500	2			CC
	R4RGC3B	400	1			CC
	Total		21		1	
Sunflower	R4RGC1	330	—[c]			TIR/CC
	R4RGC2	550	5	1		CC
	R4RGC3A	500	4		1	CC
	R4RGC3B	No band	0			CC
	Total		9	1	1	
Lettuce	R4RGC1	330	18			TIR/CC
	R4RGC2	550	6[d]			CC
	R4RGC3A	500	1[d]			CC
	R4RGC3B	400	1[d]			CC
	Total		20[d]			
Asteraceae total			50[e]	1	2	

[a] All amplifications used the F4RGC1 primer as the forward primer.

[b] Three chicory sequences were identical to each other for >97% of their nucleotides; these are represented by a single consensus sequence in the phylogenetic analyses of Figs. 1 and 2 based on amino acid sequences of conceptual translations.

[c] Sunflower amplifications using the R4RGC1 primer have previously been reported (Gedil et al., 2001) and obtained from GenBank, so they were not performed.

[d] Lettuce amplifications were performed as positive controls. Only two sequences shared less than 97% identical nucleotides to previously identified sequences. These sequences were considered novel.

[e] Total does not include positive controls matching previously isolated sequences.

amplified sequences of both groups based on the analysis of the internal motifs of the amplicons. The R4RGC2, 3A, and 3B primers amplified only the CC class as expected (Table 2). Results for the three species were identical in this respect. BLAST searches also indicated that primer sets generally isolated similar sets of sequences, with lettuce and sunflower RGCs maintained in the NR database of NCBI as the most significant matches. We did identify six novel lettuce sequences (LSRGCAA01, LSRGCAA03, LSRGCAB01, LSRGCAB02, LSRGCAC03, and LSRGCAC07), with higher similarity to RGCs from other species then to any previously identified species of the Asteraceae.

3.2. Phylogenetic analysis of asteraceae RGCs

A phylogenetic approach was taken to identify subfamilies of RGCs shared among species of the Asteraceae and to establish support for distinct groups of sequences. Phylogenetic trees consisting of newly isolated sequences and those available on Genbank were created by the neighbor-joining method as described. The Asteraceae RGC phylogeny (Fig. 1) includes data from the P-loop to RNBS-B motif for a consistent analysis among sequences that were generated using differing motifs for amplification. Three chicory sequences were highly similar with greater than 97%

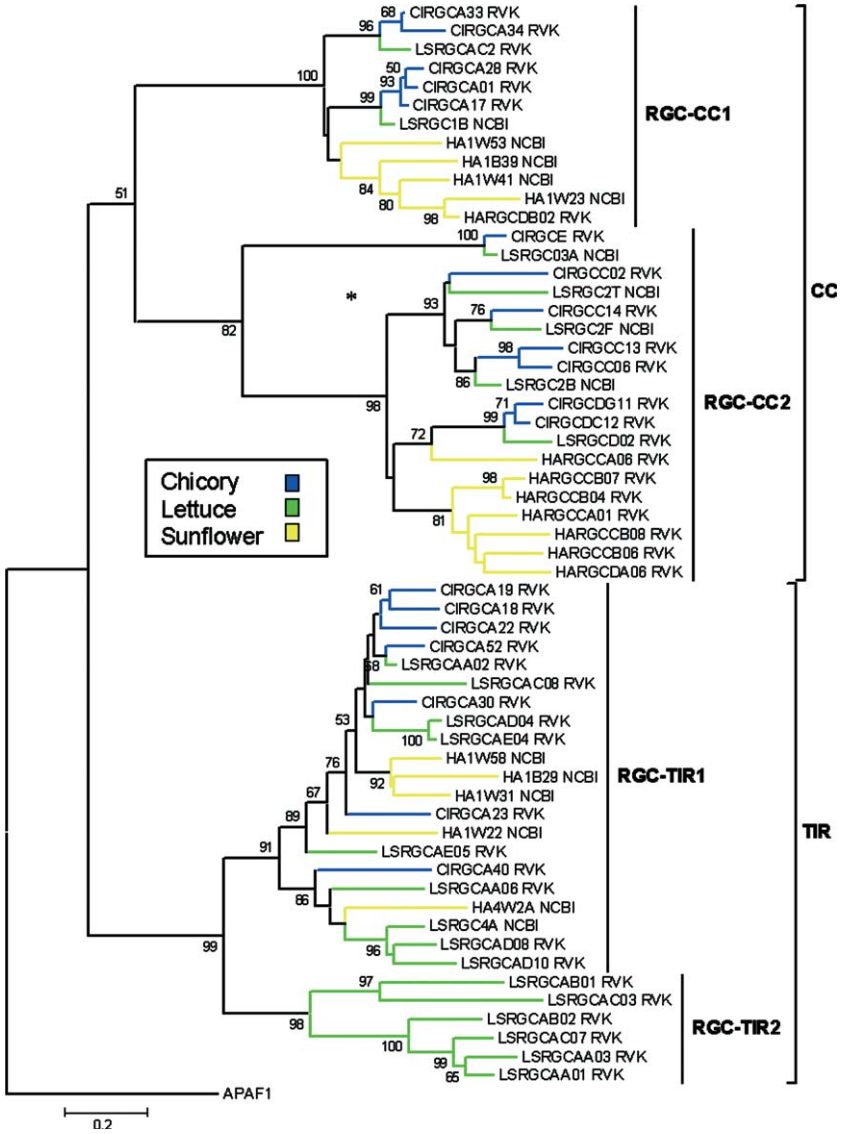


Fig. 1. Phylogenetic comparison of Asteraceae RGCs. Neighbor-joining tree consisting of 22 lettuce, 19 chicory, and 17 sunflower sequences outgroup rooted by human APAF1 as described in Section 2. Clades are labeled as defined by an arbitrary genetic distance of ~0.6. CC and TIR groups are also distinguished. The clade indicated by a star contains additional chicory and lettuce sequences that are highly similar at the nucleotide level which are not shown here. Numbers on the branches indicate the percentage of 1000 bootstrap replicates that support the node with only values of >50% reported.

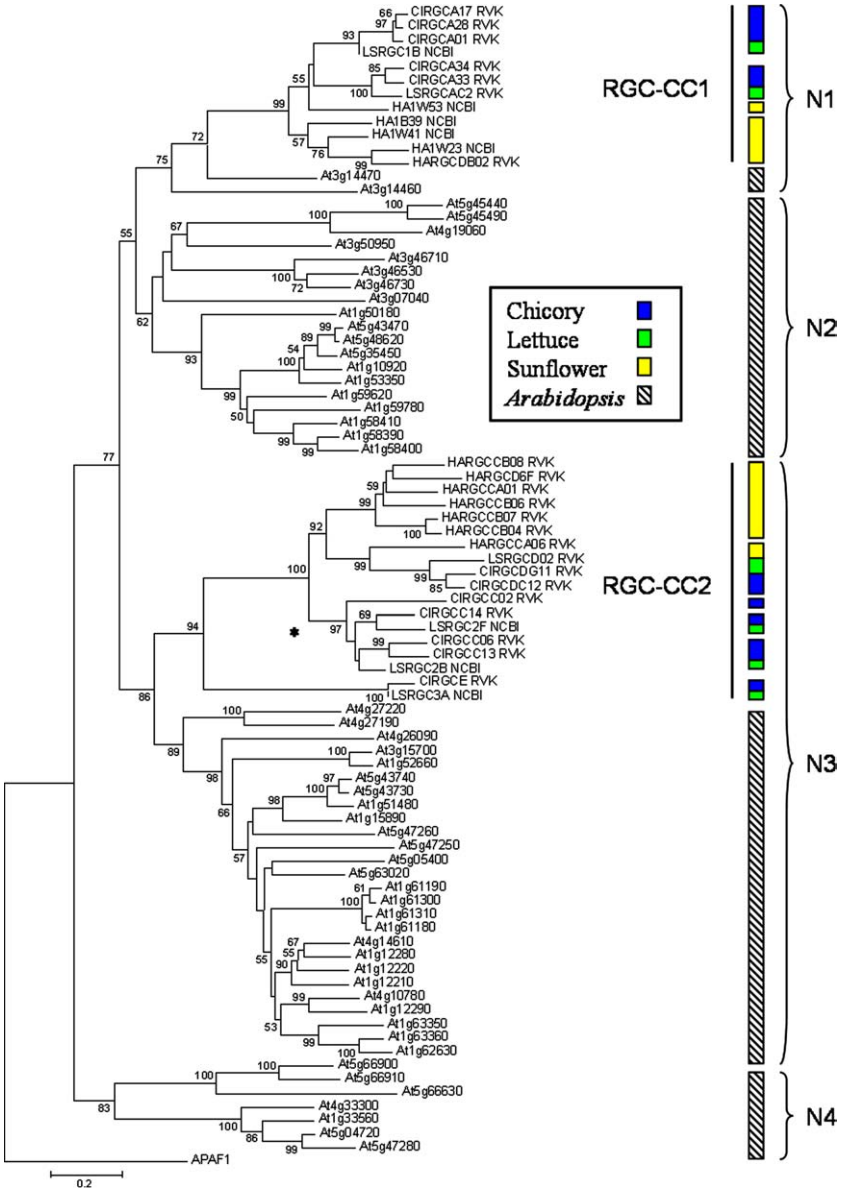


Fig. 2. Phylogenetic comparison of CC Asteraceae RGCs to *Arabidopsis*. The tree consists of 30 Asteraceae and 54 *Arabidopsis* CC NBS protein sequences outgrouped by human APAF1 as described in Section 2. Sidebars indicate ancient CC clades N1-4 as defined by Cannon et al. (2002). Not all chicory and lettuce sequences are shown and a star designates their placement. The scale represents the average number of substitutions per site. Numbers on the branches indicate the percentage of 1000 bootstrap replicates that support the node with only values of >50% reported.

nucleotide identity (noted by superscript c in Table 2) to each other and are represented by a single amino acid consensus sequence in the phylogenetic analyses (Figs. 1 and 2).

Two major clades representing CC and TIR sequences are distinguished (Fig. 1), a pattern previously described (Meyers et al., 1999; Pan et al., 2000b). Further, at a genetic distance of ~0.6, the CC and TIR groups are each subdivided into two distinct groups (termed RGC-CC1, RGC-CC2, RGC-TIR-1, and RGC-TIR-2) well supported by bootstrap analysis, and separated by long branch lengths. Representatives from each species contribute multiple sequences to three of four groups. While the bootstrap support is generally high among the branches of the RGC-CC clade, the RGC-TIR clade often lacks this support for internal

Fig. 4. Phylogenetic comparison of CC-RGC2 sub-clade. The neighbor-joining tree consisting of 24 lettuce, 8 chicory, and 1 sunflower sequence was calculated by Kimura's 2-parameter model as described. Underlined taxa are positive controls for RGC2. All sequences with the nomenclature LSRGC2 (A-W) were obtained from NCBI and are linked to *Dm3* unless otherwise labeled. LSRGC2P was not included because the sequence contains ambiguities and RGC2R and RGC2X were unavailable for analysis. The scale represents the average number of nucleotide substitution indicated by branch lengths and lengths greater than 0.05 are shown above the branch.

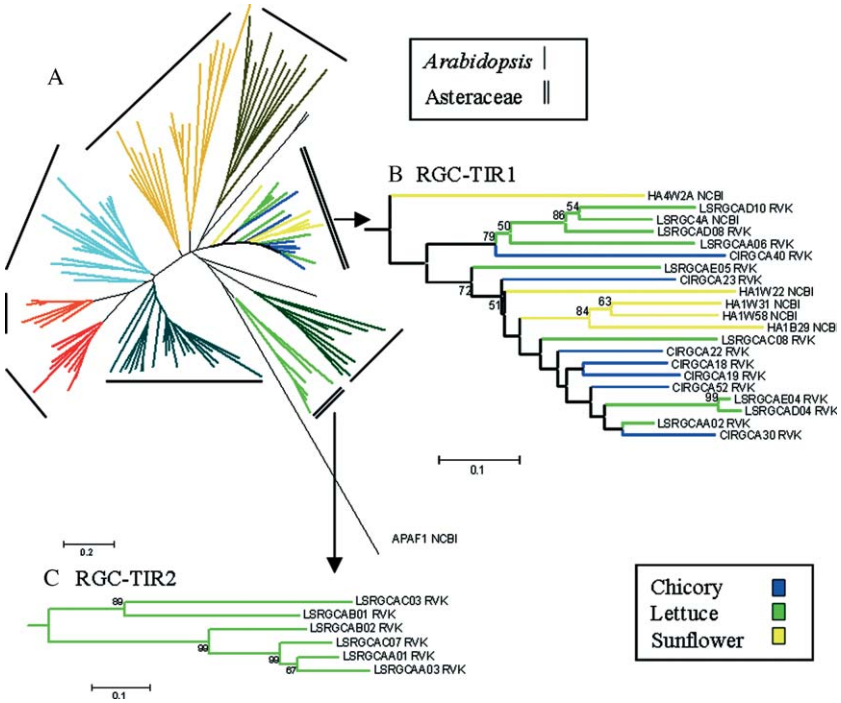
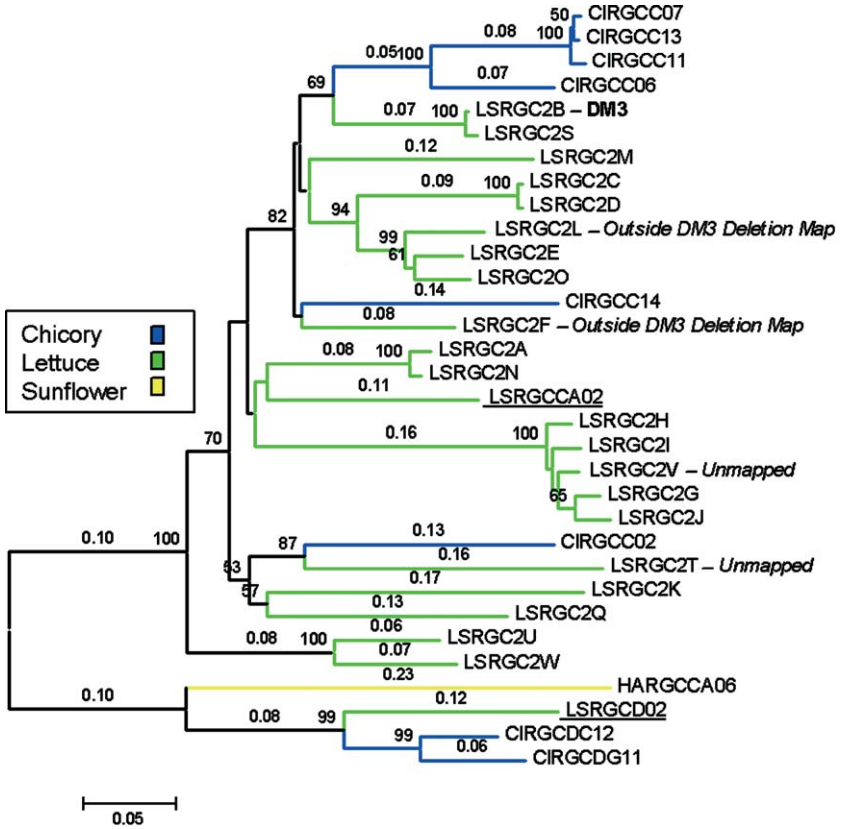


Fig. 3. Phylogenetic comparison of TIR Asteraceae RGCs to *Arabidopsis*. (A) Phylogenetic comparison of Asteraceae RGC and *Arabidopsis* NBS sequences incorporating data spanning the p-loop to RNBS-B motif. The tree consists of 27 Asteraceae and 106 *Arabidopsis* TIR NBS sequences outgrouped by human APAF1 as described in Section 2. *Arabidopsis* clades are bracketed by single black lines with colors corresponding to the seven TIR clades identified by (Richly et al., 2002). Asteraceae clades are bracketed by double black lines and arrows point to expanded subtrees of these clades. (B) RGC-TIR1 subfamily. (C) RGC-TIR2 subfamily. The color coding for the Asteraceae clades is noted in the inset. Numbers on the branches indicate the percentage of 1000 bootstrap replicates that support the node with only values of >50% reported. Acronyms following sequences indicate the source obtained, whether NCBI or previously unpublished from our lab (RVK).



branches. Two sister groups of RGC-TIR1 do however, have good support (bootstraps of 89% and 86%) and contain members from each species. The RGC-TIR2 clade with only representatives from lettuce (Fig. 1) appears to have a very different evolutionary history, though sampling bias during PCR amplifications can not be excluded. These data suggest that the last common ancestor of sunflower, lettuce, and chicory contained distinct NBS-LRR R-gene subfamilies of NBS sequences and that extant descendants remain in the genomes of the Asteraceae.

3.3. Comparison of RGCs from Asteraceae and A. thaliana

To extend the comparison of these identified gene subfamilies to a plant outside the Asteraceae, RGCs were aligned to an *A. thaliana* NBS alignment provided by R. Michelmore, A. Kozic and B. Meyers, UC Davis. *Arabidopsis*, a member of the Brassicaceae, is more distantly related to each of the Asteraceae species than they are to each other. Due to the great divergence between CC and TIR groups, separate analyses of these groups were performed. These groups were also shown to be distinct in previous phylogenetic analyses incorporating a more complete set of RGCs from taxonomically diverse plant species (Meyers et al., 1999). The phylogenies were constructed using the P-loop to RNBS-B motifs, and using the P-loop to GLPL motifs. The placement of Asteraceae clades was identical and overall topology was similar for the analyses of the two regions. The *Arabidopsis* clades were however more strongly supported in the P-loop to GLPL analysis. Also, these phylogenies were more consistent with previous studies since this is the segment commonly analyzed by others (Cannon et al., 2002; Meyers et al., 1999; Pan et al., 2000a).

The two major CC clades, RGC-CC1 and RGC-CC2 first identified in Fig. 1, are characterized by high bootstrap values (94 and 99%, respectively) separating them from *Arabidopsis* (Fig. 2). Both Asteraceae specific CC clades contain sequences from each of the three species in the family, but differ in clustering relative to *Arabidopsis*. The RGC-CC2 forms a distinct sister group to many sequences that are distributed throughout the genome of *Arabidopsis*. In contrast, the RGC-CC1 is part of a larger clade that contains two tightly linked sequences from *Arabidopsis* (At3g14460, At3g14470; Fig. 2). Thus, the analysis has identified a region of chromosome 3 in *Arabidopsis* that is potentially orthologous to the RGC-CC1 clade of the Asteraceae.

The two TIR clades from the Asteraceae (Fig. 1) remain distinct when analyzed with TIR class of RGCs from *Arabidopsis* (Fig. 3). The placement of these clades is not however, well supported by significant bootstrap values, although the tree generated was similar to those in previous studies (Cannon et al., 2002; Meyers et al., 1999; Pan et al., 2000b). The TIR-NBS clades, as identified by Richly et al. (2002), are similar, except two clades were split into paraphyletic groups in the present analysis. Within the two clades from the Asteraceae, bootstrap support is found for the external branches but is lacking for the internal nodes.

3.4. Evolution of RGC2B subfamily in the Asteraceae

The RGC-CC2 group contains LSRGC2B, which is the gene *Dm3*, conferring resistance to specific races of downy mildew (Shen et al., 2002). *Dm3* is one of 22 closely related genes, composing a large multi-gene cluster (Meyers et al., 1998). The large pool of knowledge concerning LSRGC2B paralogs in lettuce led us to investigate the relationship of homologous chicory sequences. In the RGC-CC2 group, multiple pairs of chicory and lettuce sequences form distinct groups (Fig. 2). Further, a nucleotide phylogeny containing all lettuce and chicory sequences closely related to LSRGC2B, except for LSRGC2P (excluded due to ambiguities), LSRGC2R and LSRGC2X (both unavailable), was constructed (Fig. 4). Results identify three chicory and lettuce clades among the RGC2-like sequences. These data suggest that homologs of LSRGC2B, LSRGC2F, and LSRGC2T were likely present in the common ancestor of chicory and lettuce.

4. Discussion

We have characterized members of the NBS-LRR disease resistance gene family by phylogenetic analyses of the NBS domain of homologous R-genes derived from three species of the family Asteraceae. Because the NBS domain is conserved and contains easily identifiable and therefore comparable motifs, the domain is a tractable region for the study of R-gene evolution. We have attempted to retrieve related sets of NBS sequences using four primer sets designed to amplify members of coil-coiled (CC) and toll-interleukin receptor (TIR) families of NBS resistance genes. With these and supplementary data from public databases, we have been able to investigate R-gene evolution across multiple taxonomic levels with a focus on the plant family Asteraceae. Analyses suggest that closely related genera contain strikingly similar arrays of R-genes, which can be grouped into distinct clades. Further, members of these gene families are distinct from *Arabidopsis* genes which suggests relatively recent R-gene expansions within the plant families. Taken together, these data suggest that the NBS domains of disease resistance genes may be slowly mutating, but undergoing rapid gene duplication and loss.

4.1. NBS-LRR R-gene evolution within a plant family

RGCs from the Asteraceae formed two major clades separated by long-branch lengths corresponding to CC and TIR groups of R-genes. Within these clades, further subdivisions containing members from multiple species were identified, a pattern suggesting that homologs for each of these subdivisions were present in the early ancestors of the Asteraceae. For example, one well-supported group in RGC-CC2 is composed of lettuce and chicory sequences (LSRGCD02, CIRGCDC12, and CIRGDG11) with the sunflower sequence (HARGCCA06) forming a sister branch (Fig. 1; 2). These likely represent orthologs of a sequence present in an early progenitor of the Asteraceae.

Other clades appear more dynamic with recent diversification and loss. For example, CIRGCE and LSRGC03A amplified from identical primer sets, together form the deepest branch in RGC-CC2 and appear orthologous, but lack a corresponding sunflower ortholog (Fig. 2). A second set of primers designed to capture sequences in this clade retrieved the same sequences from lettuce and chicory but again failed to amplify related sequences in sunflower. Since there are known to be few copies of this subfamily of RGCs in lettuce (Shen et al., 1998), gene loss or rapid divergence are feasible explanations for it remaining undetectable in sunflower, though sampling bias could also create this pattern.

Elsewhere, the presence of distinct species specific clades suggests that independent gene duplications have also occurred in all species (e.g., RGC-TIR2 and several smaller clades within each of the other subdivisions). Frequent gene conversion events, which would serve to homogenize genes within a genome, may provide an alternative explanation. However, if gene conversion occurs, it must not occur globally, since distinct, ancient clades exit in both the CC and TIR groups. Given the diversity of NBS sequences, concerted evolution would only seem to have a role within the most closely related subfamilies of genes. R-genes are often arranged in large multi-gene clusters, which could facilitate gene conversion events. Within the lettuce *Dm3* cluster however, analyses of nine full-length sequences revealed only one putative gene conversion event (Meyers et al., 1998), indicating that the role of concerted evolution within this cluster is not great. In *Arabidopsis*, three-fourths of R-gene clusters are composed of closely related sequences; however, the remaining clusters are composed of sequences that are not of recent common descent, indicating that R-gene diversity persists despite genomic arrangement. Indeed, the frequency of such complex R-gene clusters may suggest positive selection for R-gene cluster diversity (Richly et al., 2002). Gene duplications and cluster expansions may therefore better explain the presence of multiple, closely related sequences within a species.

Gene duplications may provide raw material for the evolution of new resistance specificities. Closely related R-genes have been shown to be experiencing diversifying selection at the LRR regions (Michelmore and Meyers, 1998). Frequent gene duplication in conjunction with diversifying selection may allow hosts to keep pace with pathogen evolution. Gene loss, on the other hand, may be frequent if R-genes carry a cost in the absence of pathogen selection. For instance, *Arabidopsis* is polymorphic for presence/absence of the RPS5 gene. The polymorphism is ancient, suggesting that it is maintained by balancing selection (Tian et al., 2002). The deletion may become fixed in the absence of the pathogen.

4.2. Comparison to Arabidopsis reveals patterns of R-gene evolution at the family level

Arabidopsis, a member of the Brassicaceae, is the model system for genomic comparisons among dicots. Differences between *Arabidopsis* and other genomes can be inferred to represent true gene loss/expansion events in *Arabidopsis*, since sampling bias from the complete draft of the *Arabidopsis* genome is negligible. As expected, *Arabidopsis* sequences generally formed distinct clades separated by greater genetic distances than comparison among species of the Asteraceae (Figs. 1 and 2). The presence of species and family specific clades suggests a radiation of R-genes in the family since the last common ancestor of the dicots. These gene family expansions are comparable to those reported for defense related genes in vertebrates (Mouse Genome Sequencing Consortium, 2002). Our observations in the Asteraceae are also similar to those described in Solanaceae and may be typical of other plant families (Pan et al., 2000a).

Our data support the conclusion that NBS-LRR genes are members of phylogenetically distinct and ancient families, members of which were present in the common ancestor of dicots (Cannon et al., 2002). The RGC-CC1 and RGC-CC2 subfamilies of Asteraceae fall within the N1.1 and N3 clades, respectively, identified by Cannon et al. (2002). As reported in Fabaceae (Cannon et al., 2002), sequence representation of the four ancient RGC-CC clades may differ among plant families; copy number may be over-represented or under-represented. For instance, *Arabidopsis* has only two copies (At3g14660 and At3g14670) in the N1.1 clade, but the species of the Asteraceae have many. From this study alone, six chicory, two lettuce, and four sunflower sequences were retrieved; a clear minimum number given the sampling strategy used. Indeed, hybridization studies estimated that approximately 10 genes closely related to LSRGC1B of the RGC-CC1 group are present in the lettuce genome (Shen et al., 1998). For a N3 ancient subclass of RGC-CC genes, both *Arabidopsis* and species of the Asteraceae are well represented, but

sequences from the Fabaceae are under-represented (Cannon et al., 2002). At least 24 genes closely related to LSRGC2B of this clade are present in the lettuce genome, most of which have been mapped to a single multi-gene cluster (Meyers et al., 1998). Finally, we did not detect any members of the N2 and N4 subclasses in species of Asteraceae, but this was not unexpected since we did not specifically target all classes of RGCs.

While phylogenetic analyses reveal evidence that NBS-LRR resistance genes were present in the last common ancestor of flowering plants, large differences in the genomic composition of these genes exist among plant families. Frequent gene duplications are a likely explanation and this pattern is consistent with the idea that evolution in these gene families proceeds by a birth-and-death process (Michelmore and Meyers, 1998; Nei et al., 1997).

4.3. Prospects for syntenic analysis

With the isolation of closely related homologs from closely related species, questions concerning syntenic conservation of R-gene clusters may be examined. LSRGC2B, *Dm3*, is the only phenotypically defined R-gene in lettuce and confers resistance to downy mildew. The *Dm3* cluster in lettuce is highly duplicated and contains at least 22 paralogs from the RGC-CC2 subfamily (Meyers et al., 1998). We detect four chicory sequences that group closely with LSRGC2B, three of which are >97% identical. Thus, recent duplications have also likely occurred in chicory. These four sequences from chicory (CIRGCC06, CIRGCC07, CIRGCC13, and CIRGCC11) are more closely related to *Dm3* (LSRG C2B) of lettuce than to other chicory sequences (CIRG CC14 and CIRGCC02) in the same clade. These latter two chicory sequences appear more similar to LSRGC2F and LSRGC2T (respectively; Fig. 4).

Interestingly, LSRGC2F and LSRGC2T are not physically located within the boundaries of the largest RGC2 cluster that contains *Dm3* along with 19 other homologs (Meyers et al., 1998). Thus, there seems to be a correspondence between the physical position of a gene in the lettuce RGC2 family and the evolutionary history of the gene; at least some of the physically displaced genes form clades with separate sets of chicory genes.

LSRGC2F and LSRGC2T differ from their closest chicory homologs by an average distance of 0.126 substitutions per site. The most closely related pair of chicory and lettuce sequences in the RGC-CC2 group (CIRGCE and LSRGC3A) is separated by smaller branch lengths of 0.04 substitutions per site (data not shown). Therefore, a conservative genetic distance cutoff of 0.05 suggests that the common ancestor of chicory and lettuce had at least 10 RGC2 copies and that lettuce paralogs have doubled recently.

Breeding systems may have a profound impact on the evolution and synteny of R-gene clusters. The *Dm3* cluster in lettuce was shown to be meiotically unstable when homozygous (Chin et al., 2001). The enhanced homozygosity in inbreeding lettuce relative to outbreeding chicory could be partly responsible for the apparent copy number increase in the lettuce lineage.

No sunflower sequences belonging to this group of genes closely related to *Dm3* were detected. Thus, while the sequences in the small clade of four chicory sequences are likely candidates for orthologs of the large RGC2 group of lettuce, syntenic positioning of the *Dm3* cluster may be difficult in sunflower. It is however, difficult to predict the likelihood of R-gene synteny as such studies have produced differing results. Within the dicot family, Solanaceae, syntenic positioning of related NBS sequences was determined (Pan et al., 2000b). However, in cereal monocot genomes, rapid reorganization of clusters was reported (Leister et al., 1998; Pan et al., 2000a). Further research into this area, including data from other plant families, is therefore warranted.

Dm3 (LSRGC2B) is necessary and sufficient for resistance to specific races of downy mildew (Shen et al., 2002). This leads to questions concerning the function of close homologs to LSRGC2B. Are any of the chicory sequences homologous to LSRGC2B responsible for resistance to downy mildew? Have new resistance specificities developed? Analyses have suggested that R-gene clusters may show syntenic positioning, but functional properties such as specificity may or may not be the same (Grube et al., 2000). The closely related homologs of LSRGC2B in chicory may provide a system to test hypotheses concerning the evolution of resistance specificity.

Acknowledgments

We thank Drs. Richard Michelmore and Steven Knapp for supplying seed and DNA samples and Dr. Ed Ryder for supplying seed samples. In addition we thank Dean Lavelle, Dr. Richard Michelmore, Dr. Steven Knapp and Dr. Ying Tan for advice at various stages of this project and Jamie Jones for editorial assistance. This work was supported in part by the National Science Foundation Research Experience for Undergraduates (REU) program, by a National Science Foundation Academic Research Infrastructure-Instrumentation grant (95-12341) and by a United States Department of Agriculture Initiative for Future Agriculture and Food Systems grant (2000-04292).

References

Altschul, S.F., Madden, T.L., Schaffer, A.A., Zhang, J., Zhang, Z., Miller, W., Lipman, D.J., 1997. Gapped BLAST and PSI-BLAST:

a new generation of protein database search programs. Nucleic Acids Res. 25, 3389–3402.

Bernatzky, R., Tanksley, S.D., 1986. Genetics of actin-related sequences in tomato. Theor. Appl. Genet., 314–321.

Cannon, S.B., Zhu, H., Baumgarten, A.M., Spangler, R., May, G., Cook, D.R., Young, N.D., 2002. Diversity, distribution, and ancient taxonomic relationships within the TIR and Non-TIR NBS-LRR resistance gene subfamilies. J. Mol. Evol. 54, 548–562.

Chin, D.B., Arroyo-Garcia, R., Ochoa, O.E., Kesseli, R.V., Lavelle, D.O., Michelmore, R.W., 2001. Recombination and Spontaneous mutation at the major cluster of resistance genes in Lettuce (*Lactuca sativa*). Genetics.

Dangl, J.L., Jones, J.D., 2001. Plant pathogens and integrated defence responses to infection. Nature 411, 826–833.

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

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

Gedil, M.A., Slabaugh, M.B., Berry, S., Johnson, R., Michelmore, R., Miller, J., Gulya, T., Knapp, S.J., 2001. Candidate disease resistance genes in sunflower cloned using conserved nucleotide-binding site motifs: genetic mapping and linkage to the downy mildew resistance gene Pl1. Genome 44, 205–212.

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

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

Kanazin, V., Marek, L.F., Shoemaker, R.C., 1996. Resistance gene analogs are conserved and clustered in soybean. Proc. Natl. Acad. Sci. USA 93, 11746–11750.

Kimura, M., 1983. The neutral theory of molecular evolution. Cambridge University Press, Cambridge.

Kumar, S., Tamura, K., Jakobsen, I.B., Nei, M., 2001. MEGA2: molecular evolutionary genetics analysis software. Bioinformatics 17, 1244–1245.

Lam, E., Kato, N., Lawton, M., 2001. Programmed cell death, mitochondria and the plant hypersensitive response. Nature 411, 848–853.

Leister, D., Ballvora, A., Salamini, F., 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., Schulze-Lefert, P., 1998. Rapid reorganization of resistance gene homologues in cereal genomes. Proc. Natl. Acad. Sci. USA 95, 370–375.

Meyers, B.C., Chin, D.B., Shen, K.A., Sivaramakrishnan, S., Lavelle, D.O., Zhang, Z., 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., 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.W., 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.

Mouse Genome Sequencing Consortium, 2002. Initial sequencing and comparative analysis of the mouse genome. Nature 420, 520–562.

Nei, M., Gu, X., Sitnikova, T., 1997. Evolution by the birth-and-death process in multigene families of the vertebrate immune system. Proc. Natl. Acad. Sci. USA 94, 7799–7806.

Pan, Q., Liu, Y.S., Budai-Hadrian, O., Sela, M., Carmel-Goren, L., Zamir, D., Fluhr, R., 2000a. 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.

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

Richly, E., Kurth, J., Leister, D., 2002. Mode of amplification and reorganization of resistance genes during recent *Arabidopsis thaliana* evolution. Mol. Biol. Evol. 19, 76–84.

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

Shen, K.A., Chin, D.B., Arroyo-Garcia, R., Ochoa, O.E., Lavelle, D.O., Wroblewski, T., Meyers, B.C., Michelmore, R.W., 2002. Dm3 is one member of a large constitutively expressed family of nucleotide binding site-leucine-rich repeat encoding genes. Mol. Plant Microbe. Interact. 15, 251–261.

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

Tameling, W.I., Elzinga, S.D., Darmin, P.S., Vossen, J.H., Takken, F.L., Haring, M.A., Cornelissen, B.J., 2002. The tomato R gene products I-2 and MI-1 are functional ATP binding proteins with ATPase activity. Plant Cell 14, 2929–2939.

Thompson, J.D., Gibson, T.J., Plewniak, F., Jeanmougin, F., Higgins, D.G., 1997. The CLUSTAL_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. Nucleic Acids Res. 25, 4876–4882.

Tian, D., Araki, H., Stahl, E., Bergelson, J., Kreitman, M., 2002. Signature of balancing selection in Arabidopsis. Proc. Natl. Acad. Sci USA 99, 11525–11530.

van der Biezen, E.A., Jones, J.D., 1998. The NB-ARC domain: a novel signalling motif shared by plant resistance gene products and regulators of cell death in animals. Curr. Biol. 8, R226–R227.

Yu, Y.G., Buss, G.R., Maroof, M.A., 1996. Isolation of a superfamily of candidate disease-resistance genes in soybean based on a conserved nucleotide-binding site. Proc. Natl. Acad. Sci. USA 93, 11751–11756.