Sentinels of Disease. Plant Resistance Genes

Robert Fluhr
Department of Plant Sciences, Weizmann Institute of Science, Rehovot 76100, Israel

Act 2, SCENE I. King Henry VI, William Shakespeare
(Enter a Sergeant with two Sentinels)
Sergeant speaks:
“Sirs! take your places and be vigilant:
If any noise or soldier you perceive
Near to the walls, by some apparent sign
Let us have knowledge at the court of guard.”

CONCEPTS IN IMMUNITY

Successful defense against an enemy requires perception of his whereabouts. In the last few years, much progress has been made in delineating the plant molecular sentinels that participate in pathogen identification. This ability is encoded by genetically “hard-wired” information, and is called “innate immunity”. It draws its origins from a phylogenetically ancient form of immunity that is common to all Metazoa and Viridiplantae. However, the rapid evolution of plant innate immunity genes has led to massive gene diversification. The appreciation of this diversification offers challenging prospects in understanding the forces that have shaped multicellular innate immunity and underlines the necessity of furthering our understanding by examining multiple plant systems.

A basic concept of innate immunity is the recognition of “non-self” components accomplished by constitutive pattern recognition receptors. Like sentinels, they alert the organism to activate defense genes. The agents that plant sentinels detect are called avirulence factors. They are produced by the pathogen and probably play a role in the process of pathogen colonization of the host. In vertebrates, non-self recognition is mediated by both innate and acquired immunity systems. Animal innate immunity consists of a limited number of pattern recognition genes that assist and prime the more complex acquired immunity response. In contrast to innate immunity, acquired immune system consists of a huge array of somatically produced recombinant proteins. Upon pathogen encounter and by the process of clonal expansion, a particular subset of recognition molecules is called into service. As acquired immunity attains its complexity during the organism’s somatic life, it offers the benefit of very complete non-self coverage (approximately $10^7$ different gene products/Mbp coding sequence) without a huge penalty in storage of genetic information (Fig. 1). Innate systems are inherently more costly in their “information content” (approximately $2 \times 10^9$ different gene products/Mbp coding sequence). However, the developmental strategy of plants, with their rigid cellular structures and absence of a circulatory system, would seem to preclude potential evolutionary development of the sophisticated acquired pattern recognition systems. Instead, plant sentinels are required to be completely cell autonomous. There is no reason to surmise that plants face a smaller number of potential pathogens than their animal counterparts. How then does an organism that relies solely on innate immunity cope? Plants have met the challenge by choosing as their sentinels a versatile group of pattern recognition molecules called resistance genes (R-genes), by greatly amplifying their numbers and by positioning them in the genome in a manner that facilitates their rapid evolution.

SENTINEL BUILDING BLOCKS

Sentinel receptor units contain a pattern recognition domain combined with accessory domains that participate in signal relay (Fig. 2). Leu-rich repeat (LRR) elements have generally been implicated in the pattern recognition role, but coiled coil (CC) and kinase domains may also be involved. Signal relay elements include membrane anchor or transmembrane domains, Toll and interleukin-1 receptor-like domains (TIR), and nucleotide binding domains (NBD). Additionally, there are animal-specific domains including caspase-activating recruitment domains and plant-specific domains that involve CC and kinase domains. Innate sentinel molecules are built up in different organisms by the combinatorial use of these components, as shown in Figure 2. Thus, evolution appears to have freely juggled with the building blocks of innate immunity, and as such they can be presented as independent motif entities.

LRR DOMAIN

Structure Function Relationships

What are the biophysical properties of the LRR domain that favored its choice as the pattern receptor for sentinels? LRR structures mediate protein-protein interaction and are the major determinants of recog-
tion specificity. The protein structure solved by the crystal structure of the porcine ribonuclease inhibitor (RI) has served as a rough structural platform for conceptualizing what the distantly related plant LRR may look like (Kobe and Deisenhofer, 1994). The 15 LRRs of RI are composed of an inner “solvent-exposed” surface rim comprising β-sheets connected by an outer rim of α-helical segments. In RI, the β-sheets are stabilized by a ladder of hydrogen bonds between conserved cysteines and Asn side chains. The α-helical segments force a curvature on the molecule so that it comes nearly full circle leaving a 60° opening. It is through this opening that the ribonuclease interacts with the solvent surface. Plant LRR can contain many fewer repeats; for example, seven as found in the sugar beet (Beta vulgaris) nematode Hs1R-pro1 R-gene (Cai et al., 1997), or many more than 15 repeats, which is usually the case. Overabundant LRR repeats in a domain would seemingly force complete circle closure. However, the plant α-helical regions tend to be much less conserved (they are shorter or nonexistent). The LRR domain structure must therefore deviate from RI in a manner that would generate a less constrained flexible domain stabilized by the hydrogen bonds between the β-sheets.

Specificity between LRR domains and their potential ligands has been inferred in animal TIR-LRR by the finding that Toll-Like Receptor 4 (TLR4) immu-

![Figure 1](image1.png)

**Figure 1.** Strategies of self and non-self recognition in innate and adaptive immunity. The shaded boxes represent infected cells. In animals, specialized cells are sources for acquired immunity and they spread throughout the organism (arrows). The genomic information content (bp) devoted to genes of acquired immunity is represented by three recombinatorial systems as exemplified in humans by the following: T-cell receptors (TCR), major histocompatibility complex (MHC), and the different antibody classes (V). The major potential pattern recognition genes in plants are shown in Figure 2. Gene number estimates are for Arabidopsis and include NBS/LRR-type (150), LRR kinase-type (174), CI-type (30), and Pto-type (70). Estimates of genomic sequence dedicated to each gene type are 5 × 10^3 for NBS/LRR and LRR-kinase and 2 × 10^3 for the rest.

![Figure 2](image2.png)

**Figure 2.** Pattern receptors are built from domains common to plants and animals. Upper, Domains have been conceptually divided into five functional categories. A, LRR, CC, and a kinase are domains that can function as pattern receptors. B, TIR, CC and caspase-activating recruitment domains (CARD) are likely involved in signal transduction by homo- or heterodimerization with acceptor molecules. C, NBD common to Nod and plant TIR/CC-NBD may serve a nucleotide-dependent switch function. D, The transmembrane motif functions to anchor attached domains or participate in transmembrane signal transfer and may also contain endocytosis signals for signal attenuation as has been found in the tomato Ve Verticillium R-gene (Kawchuk et al., 2001). E, The kinase domain probably participates in signal transduction. Lower, Known examples of domain combinations are illustrated.

Figure 2. Pattern receptors are built from domains common to plants and animals. Upper, Domains have been conceptually divided into five functional categories. A, LRR, CC, and a kinase are domains that can function as pattern receptors. B, TIR, CC and caspase-activating recruitment domains (CARD) are likely involved in signal transduction by homo- or heterodimerization with acceptor molecules. C, NBD common to Nod and plant TIR/CC-NBD may serve a nucleotide-dependent switch function. D, The transmembrane motif functions to anchor attached domains or participate in transmembrane signal transfer and may also contain endocytosis signals for signal attenuation as has been found in the tomato Ve Verticillium R-gene (Kawchuk et al., 2001). E, The kinase domain probably participates in signal transduction. Lower, Known examples of domain combinations are illustrated.
from *M. grisea* (AVR-Pita) encodes for a prepropeptide that has the features of a metalloprotease. The processed form is then transferred by an unknown mechanism from the fungus into the plant cell.

**Evolution**

If solvent-exposed regions of the LRR domain play a role in interaction with the pathogen avirulence factor, they may display high mutability. Detecting adaptive evolution is carried out by estimating base changes needed to generate amino acid changes, i.e. comparing the number of substitutions per synonymous site (Ks) with the number of substitutions per nonsynonymous site (Ka; Li, 1993). Ks is expected to exceed Ka when mutations that generate amino acid change in a particular gene are deleterious to an organism’s fitness. However, if mutations in an area of a particular gene are advantageous for the organism, natural selection favors sequence diversification, and a Ka/Ks ratio larger than 1 will emerge. A survey of plant R-genes shows that, generally, the N-terminal CC structures as well as the NBD show low average Ka/Ks ratio. In contrast, positive selection (high Ka/Ks ratio) was detected in the predicted \( \beta \)-strand region of the LRR domain in many R-gene analogs (for review, see Bergelson et al., 2001). What dictates the elevated Ka/Ks ratio? Is it entirely due to high selection pressure applied to the basal mutation rate? Or do perhaps the solvent-exposed LRR regions also have a propensity for hypermutation, a phenomenon that has been detected for immunoglobulin genes?

**TIR DOMAINS**

**Structure Function Relationships**

TIR domains are a common link between animal and plant innate immunity (Kimbrell and Beutler, 2001). In plant NBD-LRR R-genes, the TIR motif appears at the N terminus, whereas in animals the TIR domain is at the carboxyl end of a single-pass transmembrane receptor (Fig. 1). Structural analysis of the diverged TIR domains of human TLR2 and TLR4 reveals a relatively conserved surface for binding to MyD88 (Xu et al., 2000). MyD88, a signal adapter molecule, is essential for signal transduction of the immune response. The idea of surface conservation of TIR is important because it may explain how divergent TIR-containing genes can signal through the apparently singular MyD88 adapter. Database searches have yet to reveal a plant MyD88 homolog; however, the finding that mutations in *enhanced disease susceptibility locus 1* (EDS1) compromise the articulation of distinct R-genes of the TIR-NBD-LRR class argues for signal funneling through a common intermediate (Aarts et al., 1998).

The TIR domain is essential for tobacco (*Nicotiana tabacum*) N-gene-mediated tobacco mosaic virus resistance, and amino acids that affect *Drosophila melanogaster* and human TIR-dependent signaling cause either partial or full loss of N-gene function (Dinesh-Kumar et al., 2000); for example, N-gene D46 resides in a position consistent with the conserved human TLR2 surface. Mutation D46H completely eliminates function, whereas the D46Y substitution that is the normal state for the human interleukin-1 receptor had no effect. Interestingly, partial loss-of-function mutations can act as dominant negative mutations by promoting systemic hypersensitive response in the wild-type N-gene background, a result that also argues for the involvement of sentinels in higher order complexes. Unexpectedly, the plant TIR domain may contribute to determining R-gene specificity. In flax rust resistance, swapping the TIR domains of L2 and L7 R-genes switches their specificity (Luck et al., 2000). This, together with evidence for diversifying selection in the TIR region of the flax rust R-genes, may also indicate that pattern recognition operates as a complex.

**Evolution**

Phylogenetic trees of TIR domains reveal a clear division between animal and plant taxa (Kimbrell and Beutler, 2001). However, the low identity of amino acid sequence between plant and animal TIR domains (less than 20%) obviates facile structural comparison. The animal taxa are further divided into at least two independent groups, Toll-like and Interleukin receptor-like. The plant TIR family as potentially represented in the Arabidopsis genome can likewise be divided into a few TIR phylogenetic groups. One group of over 100 genes includes TIR domains as part of NBD or as part of NBD-LRR sequences. Another group of more than 30 genes is composed of solitary TIR domains or TIR domains juxtaposed to other domains of unknown function (N. Kaplan-Levy and R. Fluhr, unpublished data).

The sequence similarity between plant and animal TIR domains suggests a common unicellular ancestor. Indeed, sensitive “SMART” searches of current databases, which use reiterative sequence alignment together with a broad definition of conserved polypeptide structural elements, have revealed distant TIR homologies in prokaryotes as well (http://smart.embl-heidelberg.de/; Schultz et al., 2000). Surprisingly, despite their obviously ancient origin, TIR domains have not been detected in any of the cereal genomes, their function apparently replaced by CC domains (Meyers et al., 1999; Pan et al., 2000b).

**CC DOMAIN**

CCs are an oligomerization motif of helical structures that are made up of bundles containing two to five helices. A typical CC structure shows a heptad...
repeat where the seven positions are labeled a through g. Residues a and d tend to be hydrophobic, and the residues at the e and g positions are charged or polar. A large subset of eudicot and cereal NBD-LRR genes can be shown to contain general CC domains in their N-terminal region with over 95% probability (Pan et al., 2000b). In this context, they may serve the function of adapter TIR-like motifs (Fig. 1). However, alternative functionality is suggested in the case of the RPW8 Arabidopsis R-gene (DA2-type, Fig. 2), responsible for broad-spectrum resistance to mildew. It contains CC domains that appear at the C-terminal of a predicted transmembrane domain or signal peptide domain (Xiao et al., 2001). In this case, the CC domain may be analogous to the function played by the LRR domain and could act directly in pattern recognition of an avirulence factor common to many mildew races. Alternatively, it may play an accessory role facilitating other pattern recognition molecules to function. RPW8 is encoded by a small gene family of only five linked members, but due to the nature of CC domains, true assessment of their potential numbers in plant genomes will require sophisticated database mining.

KINASE DOMAIN
Structure Function Relationships

The tomato resistance protein kinase encoded by Pto shows remarkable similarity to interleukin-1 receptor-associated kinase and Pelle kinases that act downstream in animal TIR-LRR-directed immune response (for review, see Cohn et al., 2001). Autophosphorylation competence and the Pto activation domain were obligatory for interaction of Pto and its cognate avirulence factor (Sessa et al., 2000). Importantly, the function of Pto requires the presence of Prf gene, an NBD-LRR-type gene (Salmeron et al., 1996). Conserved activation domains delineate numerous but distant putative Pto homologs in Arabidopsis (Fig. 1); some of them appear together with other functional domains in receptor-like kinases. In functional analogy to Pto, the Arabidopsis avrPphB-susceptible kinase is necessary for Resistance to Pseudomonas syringae 5 function but not for Resistance to P. syringae subsp. maculicola 1 or Resistance to P. syringae 2-type NBD-LRR resistance function (Swiderski and Innes, 2001).

Evolution

Sequence signatures of the Pto activation domain together with the presence of sequence insertions/deletions helped define nine Pto-like phylogenic families. The clades were made up of sequence from different plant families suggesting ancient origin for Pto in the genus Solanum (Vleeshouwers et al., 2001). Indeed, orthologous Pto genes of Lycopersicon pimpinellifolium and Lycopersicon hirsutum interact with the same avirulence factor and confer resistance (Riley and Martin, 2001). Cross-species maintenance of R-gene functionality may be due to the existence of persistent pathogens that help maintain ancient disease lineages. Interestingly, L. esculentum accessions examined contained no Pto ortholog showing that Pto kinases, like their cognate NBD-LRR, are either part of metabolically dispensable pathways or are functionally redundant.

Ser/Thr kinase domains are also found as part of A1,DE-type receptors (Fig. 2). These domains play a role in signal transduction as was shown exquisitely by switching the extracellular LRR and transmembrane domains of the Arabidopsis brassinosteroid receptor kinase BRI1 with the Ser/Thr kinase domain of the rice XA21 R-gene. The resultant transgene promoted brassinosteroid-dependent resistance signaling (He et al., 2000). Shiu and Bleecker (2001) have shown that all the kinase domains of the A1,DE-type receptors belong to a single monophyletic group. The closest eukaryotic homolog to the plant receptor kinase family was found to be the D. melanogaster kinase Pelle, similar to that found for Pto.

NBD
Structure Function Relationships

NBDs are characterized by several sequence motifs found in animal ATP- and GTP-binding proteins including the Ras superfamily and the caspase pathway-related Ced-4 and Apaf-1 animal genes (Li et al., 1997). In animals, the genes regulate the activity of proteases that can initiate apoptotic cell death. As defense mechanisms in plants include apoptotic-like hypersensitive responses, the appearance of these homologies in this context is particularly intriguing. By analogy to Ras, NBD may serve as a switch function moderating the inter- or intramolecular activity of the polypeptide. Alternative structural modeling of a subset of the NBD region revealed homology to the receiver domain of His-Asp phosphoproteins typical of prokaryotic signaling pathways (Rigden et al., 2000). This would argue for NBD participation in phosphorelay as opposed to actual nucleotide binding. Direct biochemical studies with this motif are lacking. Interestingly, the Arabidopsis RPP5 NBD domain was shown to interact in the yeast two-hybrid system with RelA/SpoT gene homologs. In Escherichia coli these genes are involved in p/PPP Gpp stress effector molecule signaling (van der Biezen et al., 2000). Whether this implies a similar function in plants is not known.

A systematic approach to elucidate regions critical for activity was carried out in the N-gene NBD region. Mutations in the conserved nucleotide binding site, e.g. glycines or Lys (G216 and K222), led to loss of N-gene action but also interfered in a negative dominant fashion when the mutant transgene was present in the normal N-gene background (Dinesh-
Kumar et al., 2000). These results are consistent with that detected in other NBD-containing proteins and point to similar mechanisms of action. For example, mutations in the conserved kinase 2 position at D301 leads to complete loss of function. The equivalent mutations of the conserved Asp in CED4 disrupt the oligomerization process that is necessary for activating downstream caspases (Yang et al., 1998).

Evolution of Species-Specific R-Gene Diversity

In contrast to the highly diverged TIR/CC and LRR domains that compose NBD-LRR resistance homologs, all NBD domains contain considerable conserved stretches of sequence homology. They can be clearly divided into two distinct groups due to their different conserved subdomains (Meyers et al., 1999; Pan et al., 2000b). One group is always found linked to resident TIR sequences in its N-terminal region. This entire supergene group is notably absent from cereal databases and cannot be amplified from cereal genomes (Pan et al., 2000b). The other group, which also contains particular subdomain signatures in the NBD, is linked to resident CCs in its N-terminal region. This group is present as a superfamily of genes in both eudicot and monocot species. As EDS1 and NDR1 were shown to modulate TIR-type or CC-type multiple R-gene action, respectively, in Arabidopsis (Aarts et al., 1998), the lack of a whole group of R-gene analogs implies bifurcation of signal transduction pathways between these two great plant classes.

The ancient origin of TIR domains indicates that cereal genomes have lost or utterly corrupted their copies. The question is how and when did this occur? TIR domains have been detected in Metazoans and Gymnosperms, suggesting that the Angiosperm progenitor plant contained TIR, as well. The loss of TIR sequences in cereals (and perhaps all monocot genomes) must have occurred during monocot and eudicot divergence over the last 100 million years. While mechanisms such as unequal crossing-over can account for elimination of small multigene families, no mechanism has been suggested for complete elimination of dispersed supergene families that are detected in the genome of today’s modern plants. It is thus likely that in the ancient angiosperm progenitor only a few germline R-genes existed similar to the lower number of TIR-LRR detected in the Metazoan genomes (Fig. 3). If the great expansion in NBD-LRR gene number occurred after eudicot/monocot division, considerable diversity should be detected. Indeed, clustering of NBD sequences by the neighbor-joining comparison method showed that Asteraceae, Brassicaceae, Lianaceae and Poaceae yielded distinct family-level sequence clades (Pan et al., 2000a). Utilizing maximum parsimony and likelihood methods plus a larger sequence database, Steven Cannon working with Nevin Young and Georgiana May has confirmed and extended these findings to detect ancient origin of some non-TIR-NBD. In their analysis, taxa originating from monocots, eudicots, and gymnosperms can appear together in the same ancient sequence clades (S. Cannon, University of Minnesota, Minneapolis, personal communication; www.tc.umn.edu/~cann0010/CannonEtAl_R_genes.html). Thus, some non-TIR-NBDs were found to predate the angiosperm-gymnosperm radiation. Importantly, in any one sequence clade, particular species could be dramatically over- or under-represented. Taken together, this argues for intense diversification of NBD-LRR and indicates that no one species will serve as a good sequence model for NBD-LRR homologs.

CLUSTERING OF R-GENES

NBD-LRR sequences can reside in large extended arrays spanning millions of bp that consist of dozens of R-gene homologs. In the Arabidopsis genome, about 77% of the detected NBD sequence appear as clusters of more than one gene (http://niblrs.ucdavis.edu). To get an idea of the global genome architecture of R-genes, random NBD analogs have been isolated and mapped by high-resolution genetic mapping in soybean (Glycine max; Kanazin et al., 1996), cereal genomes (Leister et al., 1998), and Solanaceae (Grube et al., 2000; Pan et al., 2000a; Gebhardt and Valkonen, 2001). Clustering of NBD sequences was found to be evident. Juxtaposition of sequences in the genome would serve as a ready source of new variance due to unequal recombination and gene conversion events. R-gene evolution could follow a pattern of rapid positive selection of successful stochastic combinations of R-gene/aviru-
DO R GENES SHOW GENOME SYNTENY AND FUNCTIONAL CONSERVATION?

Given the rapid evolution of R-genes due to environmental pressure, it is of interest to compare the status of these genes relative to general genome architecture. Global conservation of synteny despite speciation has been observed in Solanaceae together with some specific R-gene loss (Pan et al., 2000a). Similar loss of genes, but together with the occurrence of more frequent nonsyntenic map positions, was reported for cereal R-genes and their homologs and may indicate more rapid sequence divergence in the monocots (Leister et al., 1998). The fact that cereals completely lack the more conserved TIR-LRR-type R-genes is consistent with their rapid global rate of R-gene evolution.

It is important to keep in mind that even when syntenic clusters are maintained during speciation, there is no reason to expect functional conservation. Examination of species within the Solanaceae family can readily detect conserved structural positioning but show completely different disease specificity. For example, resistance to *Fusarium* spp. in tomato corresponds to resistance to tobacco mosaic virus in pepper on chromosome 11 and the tomato nematode *Heterodera* R-gene corresponds to a potato *Phytophthora infestans* resistance loci on chromosome 4 (Grube et al., 2000; Gebhardt and Valkonen, 2001). However, at the level of closely related species, a degree of functional conservation can be detected.

This question was approached by directly examining a series of lines carrying the complete *Lycopersicon pennellii* genome introgressed into a *L. esculentum* background. Many independent *Fusarium* spp. resistance loci were detected with varying quantitative differences between them, and two out of six loci show common evolutionary origins by appearing in the same chromosomal location in more than one species (Sela-Buurlage et al., 2001). Whether these loci represent true orthologous sequence is of interest and remains to be seen.

Figure 4. Distribution of receptor-ligand association constants and probability function N. The figure conceptualizes the discrete probability that among N receptors a specific receptor of a particular association value will be found. The figure was adopted using the reciprocal of the ψ function, which describes the average of all individual affinity distributions (Lancet et al., 1993). In the cases shown, immunoglobulin genes were set at $10^{8.5}$ M$^{-1}$ which would require a receptor repertoire of $10^4$ size. Olfactory receptors are estimated to show association constants of $10^{-3}$ for their ligands and would require a receptor repertoire of $500$ to $10^4$ size. A similar estimation of affinity constants for plant R-genes could account for plant R-gene family sizes.
For example, in higher vertebrates, only 500 to 1,000 odor receptors are enough to “sense" a great spectrum of smells. How then does our nose cope with the plethora of environmental smells? Lancet and coworkers (1993) described a probability model that predicts the repertoire size of smell receptors that would ensure proper representation of receptors with a specified affinity. At the expense of reduced receptor affinity, the repertoire of potential interactions can be economic (Fig. 4). By setting an affinity value of $10^5$ M$^{-1}$, a number that has been experimentally measured in smell receptors/ligand interactions, a repertoire of smell receptors of 300 to 1,000 in number would suffice for finding at least one receptor with the specified affinity. R-genes numbers in the plant genome are in a similar range (Figs. 1 and 4). If in an analogous manner their average affinity requirements are modest, it would explain the robustness of the plant resistance response. Indeed, the limited repertoire of R-genes in any one genome can display multiple independent loci of varying quantitative efficiency (Sela-Buurlage et al., 2001). In animals, additional higher order integration of information input from different receptors yields the final sensation of smell. The “equivalent” integration in plants could imply groups of R-genes cooperating with each other perhaps in a multicomponent sentinel network. Results of cross-immunoprecipitation experiments showing one avirulence factor interacting with more than one R-gene implies that such networks may exist (Leister and Katagiri, 2000).

Genetic approaches have drawn a complex and as yet an incomplete picture of the essence of disease sentinel function. Biochemical knowledge of what precisely makes up the sentinel complex as well as understanding the interplay of pathogen colonization and pathogenesis response presents an intriguing challenge that will see continued multidisciplinary and multisystem approaches.

Received August 21, 2001; accepted September 10, 2001.

LITERATURE CITED


Rigden DJ, Mello LV, Bertioli DJ (2000) Proteins 41: 133–143