Evolution of Plant NLRs: From Natural History to Precise Modifications

Janina Tamborski and Ksenia V. Krasileva
Department of Plant and Microbial Biology, University of California, Berkeley, California 94720, USA; email: kseniak@berkeley.edu

Abstract

Nucleotide-binding leucine-rich repeat receptors (NLRs) monitor the plant intracellular environment for signs of pathogen infection. Several mechanisms of NLR-mediated immunity arose independently across multiple species. These include the functional specialization of NLRs into sensors and helpers, the independent emergence of direct and indirect recognition within NLR subfamilies, the regulation of NLRs by small RNAs, and the formation of NLR networks. Understanding the evolutionary history of NLRs can shed light on both the origin of pathogen recognition and the common constraints on the plant immune system. Attempts to engineer disease resistance have been sparse and rarely informed by evolutionary knowledge. In this review, we discuss the evolution of NLRs, give an overview of previous engineering attempts, and propose how to use evolutionary knowledge to advance future research in the generation of novel disease-recognition capabilities.
1. INTRODUCTION

1.1. Extracellular and Intracellular Branches of Plant Immune Perception

Plant immune receptors monitor the extracellular and intracellular compartments for signs of a pathogen invasion. Numerous plasma membrane–localized receptors, commonly referred to as pattern recognition receptors, surveil the extracellular space for signs of an invading pathogen. They recognize conserved molecular patterns derived from microbial molecules (microbe-associated molecular patterns) or from the actions of microbial molecules (damage-associated molecular patterns), collectively referred to as danger signals (57). The recognition of danger signals leads to the activation of well-characterized immune responses that include but are not limited to the production of reactive oxygen species (ROS), an increase in cytosolic calcium concentration, activation of mitogen-activated protein (MAP) kinase cascades, and the expression of defense genes (18). Pattern recognition receptors, their signaling partners, and their outputs have been reviewed elsewhere (37, 75).

Cytoplasmic immune receptors recognize foreign molecules that are secreted into the plant cell by an invading pathogen as well as enzymatic modifications to plant components. These are called nucleotide-binding leucine-rich repeat receptors (NLRs) after the two core domains shared...
among the majority of these proteins. The secreted pathogen molecules are termed effectors, and their function supports invasion of the pathogen, often through the suppression of plant immunity. The recognition of effectors by NLRs leads to some of the same responses as the recognition of extracellular patterns. These include the accumulation of ROS, activation of MAP kinase cascades, and defense gene expression; however, the kinetics of these responses differs from that of surface receptors (45). In addition, NLR activation culminates in localized cell death, known as the hypersensitive response (HR).

1.2. NLRs Are Intracellular Immune Receptors

NLRs have a distinct domain architecture that consists of a nucleotide-binding (NB-ARC) domain and a series of C-terminal leucine-rich repeats (LRRs), and most have an N-terminal extension consisting of a Toll/interleukin-1 receptor (TIR) domain, a coiled-coil domain (CC), or a divergent coiled-coil domain (CC_{R}) similar to the RESISTANCE TO POWDERY MILDEW 8 (RPW8) domain (113). NLRs form three classes on the basis of the N-terminal domain and evolutionary history of the NB-ARC: TIR-NLRs (TNLs), CC-NLRs (CNLs), and RPW8-NLRs (RNLS) (Figure 1) (113). A few NLRs do not have all of the canonical domains; for instance, RESPONSE TO THE BACTERIAL TYPE III EFFECTOR PROTEIN HOPBA1 (RBA1) is a TIR-only protein that recognizes the bacterial effector HopBA1 (96), and proteins containing only TIR or RPW8 domains can elicit resistance to pathogens on their own (94, 135, 144). NLRs can directly bind and recognize effectors (e.g., 46, 62) or indirectly recognize the modification of another plant component through effector function (129, 150). Plants also have immune receptors with integrated domains (IDs) that mimic pathogen targets and are activated in response to modification by the effectors (28, 70, 105).

In general, we can divide NLRs into two functional groups: direct/indirect sensor NLRs that are involved in the recognition of invasion and helper NLRs that are genetically required by other NLRs for immune activation (65). In some cases, NLRs work in a pair consisting of a sensor and a helper and are genetically linked. In some species, helper NLRs are proposed to have arisen from a single pair before expanding into complicated networks of many NLRs, such as the NLR REQUIRED FOR CELL DEATH (NRC) gene family, which has expanded in the asterids and serves as helpers for a number of sensor CNLs (140–142). RNL helpers that function downstream of sensor NLRs and are required for both resistance and cell death include ACTIVATED DISEASE RESISTANCE 1 (ADR1) and N REQUIREMENT GENE 1 (NRG1) (25, 71, 102, 143).

1.3. Mechanisms of NLR Activation

The molecular mechanisms underlying activation of sensor NLRs and of their helpers are expected to differ significantly and are not yet fully understood. However, their shared domains and common evolutionary origin suggest that multimerization through the NB-ARC domain following exchange of ADP for ATP is a key step in NLR activation. For this reason, NLRs are often referred to as molecular switches in immune signaling (123). The ADP-bound state is thought of as the off state, in which the LRR associates with the NB-ARC domain, thereby stabilizing the NLR in the inactive state (124). The activation of NLRs is generally associated with the ATP-bound state and is referred to as the on state.

Cryo-electron microscopy structures of an indirect sensor protein, Arabidopsis thaliana HOPZ-ACTIVATED RESISTANCE 1 (ZAR1), have demonstrated both states and suggested the existence of a third, intermediate state (133, 134). These structures showed that the ADP-bound form was monomeric and had multiple intramolecular contact points between the LRR and NB-ARC.
a  Phylogenetic relationship of NLRs across monocots and dicots

b  NLR functions

<table>
<thead>
<tr>
<th>Direct binders</th>
<th>Indirect binders</th>
<th>Integrated domain</th>
<th>Helpers</th>
</tr>
</thead>
<tbody>
<tr>
<td>RPP1, MLA</td>
<td>ZAR1, RPS2, RPM1</td>
<td>RRS1, RGAS</td>
<td>RP54, RGA4, NRG1, NRC, ADR1</td>
</tr>
</tbody>
</table>

1 residue substitution per site

(Caption appears on following page)
While all TNLs to date have been found to signal through and strictly require required self-association and depended on the conserved catalytic glutamate (59, 131). Plant TIRs produced a third product, v-cADPR (131). The enzymatic activity in both systems + TIR from STERILE ALPHA AND TIR MOTIF CONTAINING 1 (SARM1), were capable through its N-terminal multiple surfaces (16, 138, 148). While the active ZAR1 resistosome associates with a membrane think that they would assume a similar wheel-shaped structure based on oligomerization of the negative inhibition. The addition of ATP induced an oligomeric state and the formation of a wheel-shaped pentamer called the resistosome (133), reminiscent of the mammalian apoptosis (151) that represents the active state of Apaf-1 and the inflammasomes that assemble upon activation of mammalian NLRs (126). The CC of ZAR1 contributes to its oligomerization by forming an α-helical barrel (133). The active complex associates with the plasma membrane, and the charged residues inside the funnel are required for the initiation of cell death and disease resistance (133).

How ligand binding or modification affects the nucleotide state of the NB-ARC domain needs further elucidation. One possibility, based on observations from numerous studies, is that the binding of an effector leads to conformational changes within the NLR molecule, enabling exposure of the NB-ARC domain and thereby facilitating the exchange from ADP to ATP (86, 125, 137). This is also the case in ZAR1, in which binding of the uridylated host protein kinase PBS1-LIKE 2 (PBL2) to the ZAR1/RKS1 (RESISTANCE-RELATED KINASE 1) complex resulted in the rotation in the NB-ARC relative to the other domains and a loss of affinity to ADP (134). However, in a different model, known as the equilibrium-based switch model, a pathogen effector showed no significant binding affinity to the ADP-bound state of the NLR (15). In this model, the NLRs constitutively cycle between the on and off states, and the pathogen effector stabilizes the on state and shifts the equilibrium in favor of the on state, leading to activation of defense responses (15). Both models probably coexist and are utilized by different NLRs. Allowing different modes of activation would not only make immune responses more robust against pathogen interference but would also allow adaptation to varying effector binding affinities.

To date, no structures of active TNL complexes have been reported; however, it is logical to think that they would assume a similar wheel-shaped structure based on oligomerization of the NB-ARC domain. TNLs are known to form oligomers, and TIR domains can self-associate across multiple surfaces (16, 138, 148). While the active ZAR1 resistosome associates with a membrane through its N-terminal α-helix, the active TNL complex likely signals through the enzymatic breakdown of NAD+ by the TIR domains (59, 131). Multiple plant TIRs, as well as a mammalian TIR from STERILE ALPHA AND TIR MOTIF CONTAINING 1 (SARM1), were capable of cleaving NAD+ into nicotinamide and cyclic ADP-ribose (cADPR) (59, 131); however, some plant TIRs produced a third product, v-cADPR (131). The enzymatic activity in both systems required self-association and depended on the conserved catalytic glutamate (59, 131). Activation of NLRs induces downstream responses that can ultimately lead to HR, a form of localized cell death. While all TNLs to date have been found to signal through and strictly require the ENHANCED DISEASE SUSCEPTIBILITY 1 (EDS1) and PHYTOALEXIN DEFICIENT 4

**Figure 1** (Figure appears on preceding page) NLR evolution and functions in flowering plants. (a) Maximum likelihood phylogeny of 7,133 NLRs from 11 dicots (green), 7 monocots (yellow), and 1 moss (red) based on NB-ARC domain alignment (11). Major classes of NLRs are depicted as arcs: RNLs (light purple), TNLs (medium purple), and CNLs (dark purple). Examples of well-characterized NLRs with different functions are marked on the tree in blue text (sensors) and gray text (helpers). The tree is rooted on the longest internal branch, and is based on the NB-ARC domain. Bootstrap values >80 are indicated on the tree as black circles. Monocots (Poales) are Oryza sativa, Sorghum bicolor, Triticum aestivum, Setaria italica, Hordeum vulgare, Zea mays, Brachypodium distachyon. Dicots are Arabidopsis thaliana, Medicago truncatula, Solanum tuberosum, Solanum lycopersicum, Vitis vinifera, Glycine max, Malus domestica, Prunus persica, Eucalyptus grandis, Fragaria vesca, Populus trichocarpa. Moss is Physcomitrella patens. An interactive version of this tree constructed by Dr. Paul Bailey following the methodology described in Bailey et al. (10), is available at [https://itol.embl.de/tree/149155221225191791507546488](https://itol.embl.de/tree/149155221225191791507546488). (b) Schematic diagram of NLR functions. Blue proteins indicate sensors, including direct binders, indirect binders, and NLRs with integrated domains; gray proteins are helpers. Pathogen-derived effectors are shown as circles; plant proteins targeted by the pathogen, as well as integrated domains, are shown as hexagons, and the effector recognition site is a triangle.
(PAD4) genes encoding for lipase-like proteins (50), exactly how the enzymatic products of the TIR domains affect downstream signaling components remains unknown. Recent studies have substantially added to our understanding of how RNLs function downstream of sensor NLRs (25, 71, 102, 143). The new data suggest that NRG1 interacts with EDS1 and SENESCENCE-ASSOCIATED GENE 101 (SAG101) and that this complex is strictly required for initiation of HR in *A. thaliana* (102). This complex might directly execute HR, as the RPW8 domain of NRG1 is reminiscent of pore-forming toxins and it may insert into the plasma membrane to disrupt membrane integrity (11). Some CNLs genetically rely on NON-RACE SPECIFIC DISEASE RESISTANCE 1 (NDR1) (1, 27), a protein that is anchored in the plasma membrane. Since the active resistosome of ZAR1 was associated with membranes, it is possible that NDR1 might be required for the suggested pore-forming activity. Together, these studies have brought us much closer to understanding the initiation and function of HR, which is a crucial step toward a better understanding of NLR-mediated immunity.

In the following sections, we discuss different evolutionary processes that yield the wide variety of NLRs, illustrate how phylogenetic and evolutionary analyses can indicate which modes of activation are deployed by which NLRs, and explain how we can apply this knowledge to engineer novel disease resistance.

## 2. GENOMIC BASIS OF NLR EVOLUTION

### 2.1. NLR Evolution Across Angiosperms

From analyses of more than 100 sequenced plant genomes, common patterns of NLR evolution have emerged that were not apparent in studies of a single species or plant family (54, 113, 149). On the basis of their NB-ARC phylogeny, the TNLs, CNLs, and RNLs form three monophyletic groups and have unique N-terminal domains that likely represent ancestral fusions of the TIR domain, CC, and CCR to an ancestral NB-ARC domain (Figure 1a). Representatives of all three NLR types are present in the basal plant lineage *Amborella*, suggesting that the split into the clades is ancient (113). Ancestral reconstruction of NLRs from 22 representative angiosperms suggested that a basal plant had approximately 23 groups of NLRs (90, 113). An orthogroup analysis of 20,571 NLRs across 75 plant genomes demonstrated that only 38 of 311 NLR families are conserved across monocots and dicots, while the other groups represent lineage-specific gene expansions (149).

The copy number of RNLs, TNLs, and CNLs varies across plants (Figure 1a). The RNL clade is usually characterized by a low copy number (54, 70, 113, 149), with the exception of gymnosperms (there are 31 ADR homologs in spruce) (145). RNLs show remarkable intron conservation, with *Amborella* and dicots sharing four introns and monocots three introns (the second intron is lost) (90, 113). The separation of NRG and ADR occurred prior to the divergence of angiosperms, and they are still conserved in syntenic blocks across flowering plants (113). The NRG genes, but not the ADR genes, have been lost in several lineages (113).

The TNLs form two subfamilies, TIR1 and TIR2 (105), with only TIR2 NLRs retained in monocots (94, 105). TIR1 NLRs have proliferated in many dicot species but are absent in some dicot lineages (113). Across all flowering plants, TNLs show remarkable conservation of intron/exon junctions: The first intron separates TIR from NB-ARC, the second separates NB-ARC from LRRs, and the third separates the first LRR from the rest of the protein (90, 113). In contrast, CNLs do not share conserved introns (90, 113), and current data suggest that ancestral CNLs were intronless and that introns in CNLs were likely gained later in evolution (113).

The CNLs are subdivided into at least four distinct groups based on amino acid motifs conserved across a large evolutionary distance (139). Two groups shared across monocots and dicots...
contain an EDVID motif, such as ZAR1 and RPM1 proteins in *Arabidopsis*, Rx in potato, and Sr33/MLA in Poaceae (139). Some CNLs also contain a functionally conserved N-terminal methionine, alanine, aspartate, alanine (MADA) motif (2) within the first α-helix, which is rearranged upon oligomerization and mediates membrane anchoring (133).

In the context of the global evolutionary history of NLRs (Figure 1a), it is evident that sensor and helper roles evolved multiple times. The NRC helper clade emerged before the split of Caryophyllales and asterids more than 100 million years ago and expanded in Solanaceae (140); yet the RNL helper clade is even older, dating to before the split of gymnosperms and angiosperms (113). Similarly, paired NLRs such as RESISTANT TO RALSTONIA SOLANACEARUM 1/RESISTANT TO PSEUDOMONAS SYRINGAE 4 (RRS1/RPS4) and R-GENE ANALOG 4/R-GENE ANALOG 5 (RGA4/RGA5) are paraplyhyetic (Figure 1a). The sensors RESISTANT TO PSEUDOMONAS SYRINGAE 2 (RPS2) and RESISTANCE TO PSEUDOMONAS SYRINGAE PV. MACULICOLA 1 (RPM1) guard the same plant protein, RPM1-INTERACTING PROTEIN 4 (RIN4), but diverged at least before the split of monocots and dicots, suggesting that their function likely arose independently. Similarly, genes encoding direct binders that recognize the same pathogen [such as RECOGNITION OF PERONOSPIRA PARASITICA 13 (RPP13) and RPP1] do not cluster together. This finding demonstrates that functional similarity between NLRs does not necessarily reflect their evolution.

2.2. Lineage-Specific Clade Expansions and Contractions

There is a 100-fold variation in the number of NLRs across plant genomes, ranging from a few dozen in papaya, kiwi, cucumber, and watermelon (8, 54, 78, 81, 149) to several thousand in hexaploid wheat (4). Even closely related species can show lineage-specific expansions and contractions (4, 8, 9, 55, 56, 63, 78, 82, 118, 149). While the selection mechanisms driving NLR expansions and contractions remain elusive, they can reflect plant lifestyle and be shaped by the selection pressures from the environment. Plant lifestyles that have been correlated with NLR history include aquatic environment (9) and either being dioecious (kiwi, papaya) or having separate male and female flowers (maize, cucumber) (8). Population genetics of NLR repertoires indicate that environmental pressure shapes NLR diversity within a single species, such as adaptation to pathogens in wild tomato (119, 120) and *Arabidopsis* (128), and can lead to long-term maintenance of polymorphisms (66).

How can the expansion or contraction of NLRs be achieved in a short evolutionary time? A burst of CNL expansions in Solanaceae and Poaceae is attributable in part to the activity of long terminal repeat (LTR) retrotransposons (67). The repetitive nature of NLRs themselves also aids in their evolution, deploying recombination for gene conversions (92, 111) and replication machinery for local duplications (72).

2.3. NLR Allelic Diversity and Novel Gene Fusions

Expansion and contraction of NLR genes can be observed even among closely related species; however, gene duplications alone are not sufficient to generate new pathogen recognition. The rapid NLR evolution proposed in the so-called birth-and-death model (91) holds true today. Current genomic data sets support the diversification of NLRs through intragenic and intergenic recombination and gene conversion that generate chimeric LRRs (111), as well as point mutations in surface-exposed regions of LRRs (83, 87). Such NLRs are capable of recognizing highly variable and even structurally unrelated effectors, most likely through a direct binding mechanism, and can provide resistance against multiple pathogens (17, 34, 83, 117). NLRs with high allelic variation can arise in either TNL or CNL clades and are not monophyletic.
More recently, NLRs with IDs emerged as sources of new recognition specificities (28, 70, 105). A specialist NLR-ID clade in Poaceae, major integration clade 1 (MIC1), continuously shuffles IDs, thereby facilitating the generation of new ID sensors (10, 121). NLR-IDs show evidence of duplications and intrachromosomal translocations, but the exact genomic mechanism of domain integration remains elusive. An analysis of NLR-IDs in wheat suggests that once the ID is integrated near the NLR in the genome, alternative splicing gives rise to a fused transcript (4).

Rapid generation of new recognition specificity can come at a cost: autoimmunity (19, 32). Several allelic variants of NLRs have been linked to autoimmunity in Arabidopsis (3, 31, 77), representing a possible mismatch of NLRs and their guardees. An allelic variant of the A. thaliana NLR RPP7 was recently shown to cause an autoimmune response when combined with incompatible alleles of RPW8 (67). One might speculate that the low number of NLRs in genomes of highly inbred crops such as maize is the result of autoimmunity leading to the loss of mismatched NLRs (8, 118).

2.4. Functional Networks Among Sensor and Helper NLRs

Flor’s (51) original gene-for-gene postulate in R gene–effector recognition evolved into a new paradigm of functional networks among both sensors and helpers. RPS2 and RPM1 are classic examples of sensor NLRs that can track multiple effectors by guarding the effector target hub RIN4 (7, 13, 39, 84, 85). Other NLRs such as ZAR1 monitor multiple paralogous guardees, each tracking a distinct effector (12, 73, 79, 108, 132). Alleles of the RRS1 encoding NLR-ID with an integrated WRKY domain can detect at least two effectors, AvrRps4 and PopP2 (42, 89, 106, 107), and the RRS1/RPS4 pair can provide resistance against three different pathogens (95). While multiple effector recognition specificities can be encoded in one NLR, independently evolved NLRs can also provide recognition of the same effector. In soybean and Arabidopsis, RIN4 is guarded by the evolutionarily unrelated NLRs Rpg1b/Rpg1r and RPM1/RPS2, respectively, suggesting that recognition of the Pseudomonas effectors AvrRpm1 and AvrB evolved independently (6). Recognition of protease activity of another Pseudomonas effector, AvrPphB, evolved at least twice in Arabidopsis and barley through NLRs guarding the same target, AVRPPHB SUSCEPTIBLE 1 (PBS1) (24). Finally, orthologous NLRs, such as MLA/Sr33/Sr50 and Rx1/Gpa2, evolved recognition of distinct effectors through allelic variation (83, 117). Altogether, these studies demonstrate that recognition specificity of a sensor NLR cannot be assigned on the basis of sequence identity alone.

Helper NLRs can form functional networks of their own. In Arabidopsis, TNLs and CNLs can signal through proteins belonging to the NRG1 and ADRI subclades, with TNLs showing a stronger preference for NRGs and CNLs for ADRs (25, 71, 102, 143). The NRC helper clade represents a more recent network, supporting a closely related CNL sensor clade (140). Whereas the protein products of paired NLR genes (such as RGA4/RGA5 and RRS1/RPS4) also show physical interaction (29, 61), functional dependence of NLRs on NRC and RNL has been shown so far only genetically (25, 71, 102, 140–143). Whether protein–protein interactions are unique to paired NLRs and have been lost during formation of helper networks remains to be determined.

3. PATTERNS OF CONVERGENT EVOLUTION IN NLR BIOLOGY

3.1. Dependence of Sensors on Helpers Arose More Than Once During NLR Evolution

The RNL clade is an ancient, conserved group of helpers shared between monocots and dicots, while the helper function of the NRCs arose independently (Figure 2a). The NRC clade most
Patterns of convergent evolution in NLR biology. (a) Sensor/helper networks and pairs have evolved independently multiple times during NLR evolution. Sensor NLRs are involved in recognition of pathogen invasion, whereas helpers are genetically required by sensors to execute immune signaling. They interact in networks, as is the case for helpers of the RNL and NRC classes (i), or in a paired relationship (ii), which is usually associated with genetic linkage between the helper and sensor in a head-to-head orientation. Representative examples of each NLR type are shown in parentheses. (b) Regulation of NLRs through small RNAs (sRNAs) is conserved across most plant lineages. The NLRs are regulated by sRNAs, including microRNAs (miRNAs) that are derived from NLRs themselves and involved in transcriptional silencing (i) and sRNAs that target transposable elements (TEs) and NLRs through RNA-dependent DNA methylation (ii). A representative example of such an miRNA is shown in parentheses. (c) The formation of a multimeric complex of activated NLRs is conserved across kingdoms. (i) The known resistosome, as demonstrated by Wang et al. (132, 133). We propose several hypothetical ones (ii, iii) that could be analogous to the variety of inflammasomes formed by mammalian NLRs. Likely emerged from a solanaceous NLR pair more than 100 million years ago and has since evolved into a functional network of helpers and sensors (140). Even more recently, two independent clades of TNLs and CNLs, exemplified by RRS1/RPS4 and RGA5/RGA4, subdivided into sensors and helpers in both Poaceae and Brassicaceae (28, 95). Intriguingly, Poaceae and Brassicaceae genes encoding sensor and helper pairs are genetically linked in head-to-head orientation, and the sensor often carries an ID (Figure 2b) (10, 95). The genomic and evolutionary mechanisms that drove convergent head-to-head orientation of gene pairs, fusion with IDs, or functional subdivision into helpers and sensors are currently unknown.
The dependence of sensor immune receptors on additional receptors is not unique to plants. In animals, the NLR FAMILY, APOPTOSIS INHIBITORY PROTEIN 5 (NAIP5) is a flagellin-binding sensor that depends on the helper NLR FAMILY, CASPASE ACTIVATION AND RECRUITMENT DOMAIN-CONTAINING 4 (NLRC4) to execute the response (126). Even outside the NLR family, ligand-binding receptor kinases such as FLAGELLIN SENSING 2 (FLS2) and BRASSINOSTEROID INSENSITIVE 1 (BRI1) that functionally act as sensors depend on the physical interaction with the coreceptor BRH1-ASSOCIATED RECEPTOR KINASE 1 (BAK1) for complex activation and downstream signaling (35, 93). Therefore, the requirement of an additional partner by ligand-sensing receptors is a general evolutionary theme. It is important to understand whether such a requirement is imposed by common functional constraints for receptors to initiate signaling cascade or whether it presents an additional level of regulation.

3.2. Regulation of NLRs by Small RNAs

Misregulation of NLRs is costly; therefore, they are tightly regulated on both the messenger RNA and protein levels (76, 115). A global analysis of NLRs in plant genomes revealed a common pattern: NLRs are regulated by microRNAs (miRNAs) that can be generated from duplicated NLRs (Figure 2b). Some miRNAs, such as miR482/2118, which targets CNLs, appeared a long time ago; these miRNA families are conserved across most plant lineages and target the conserved P-loop region of the NB-ARC domain (54). Other miRNAs appeared more recently and are lineage specific (58, 76, 80, 99, 110, 115, 145, 147, 149). Continuous evolution of new miRNAs has been linked to lineage-specific expansion of NLRs, with new miRNAs likely to be derived from inverted duplications of target NLR sequences (149).

Regulation of NLRs by miRNAs bears similarity to another group of rapidly proliferating genetic elements that are also regulated by small RNAs (sRNAs): transposable elements (TEs). Both NLRs and TEs duplicate rapidly, show lineage-specific expansions and contractions, and can jump around in the genome. Both newly inserted TEs and recently duplicated genes are regulated by epigenetic marks through RNA-dependent DNA methylation guided by sRNAs (97, 103). A global analysis of sRNAs and associated changes in A. thaliana epigenetic marks upon Pseudomonas syringae infection showed that both TEs and NLRs are derepressed early in infection and start to produce sRNAs, many of which map to both NLR and TE loci (23). While TEs are silenced later in the course of infection by the process of RNA-directed DNA methylation (RdDM) depositing de novo DNA methylation at the genomic loci, many NLRs, including ADR1, continue to be expressed, despite the continuous presence of sRNAs that can target them (23).

Coregulation of NLRs by TEs inserted into the promoter elements has an important functional role (Figure 2b). In rice, the insertion of a miniature inverted repeat transposable element into the promoter of the NLR gene PigmS constrains its expression to pollen and silences its transcription through RdDM in other plant tissues (41). Functionally, the PigmS protein is a dominant suppressor of another NLR, PigmR, which confers resistance to Magnaporthe oryzae. When PigmS is silenced, the PigmR protein is active and provides resistance in vegetative tissues; however, in pollen, it is suppressed by PigmS, thereby reducing the yield penalty for the plant of having activated immune signaling in the grain (41). A better understanding of NLR regulation by sRNAs and reversible epigenetic marks could inform new strategies to design NLR promoters that maintain the balance between activation of immunity and yield penalty.

3.3. Similarities to Mammalian Intracellular Immunity

Independent evolution of common NLR features is evident not only across plant lineages but also across kingdoms. In animals, a distant homolog of the NB-ARC domain, NACHT, has been
independently combined with LRRs and different signaling domains: CASPASE ACTIVATION AND RECRUITMENT DOMAIN (CARD), BACILLOVIRUS INHIBITOR OF APOPTOSIS PROTEIN REPEAT (BIR), PROTEIN PYRIN DOMAIN (PYD), DEATH DOMAIN (DD), and DEATH EFFECTOR DOMAIN (DED) (127). Mammalian NLRs serve a similar intracellular surveillance function and can induce cell death upon either pathogen perception or autoimmune response (64). In fungi, the NACHT domain-containing protein Het-E triggers nonself recognition and the induction of cell death, called heterokaryon incompatibility (100). Interestingly, the fungal HeLo domain with a four-helix bundle similar to RPW8 in RNLs can insert itself into the membrane, thereby forming a pore to induce cell death (38).

A recent breakthrough in the structural biology of plant NLRs has come from solving the structure of the active oligomeric complex of ZAR1 together with its guardses RKS1 and PBL2 kinase (133). The ZAR1 resistosome is structurally similar to the apoptosome of Apaf-1 (151) and the inflammasome of mammalian NLRs (Figure 2c) (104, 126, 133). While the structural similarities between the inflammasome and the resistosome are clear, many questions remain unanswered. First, is the CC of CNLs sufficient to form a pore in membranes and induce cell death, or does it require additional factors? The RNLs and their RPW8 domains, which are sufficient to create a pore in structurally similar fungal HeLo proteins, are clear candidates.

Second, do all sensors form homogeneous resistosomes, or does the composition of resistosomes vary? Is there, for example, an equal ratio of paired helpers and sensors, or does a single sensor initiate complex formation of several helpers (Figure 2e)? The latter is the case for mammalian NLRs such as NAIP5 and NLRC4 (126). A single NAIP5 sensor is activated by flagellin perception and, upon conformational change, binds to its helper NLRC4, inducing further NLRC4-NLRC4 oligomer formation (126). The final inflammasome forms an open-ring structure with a single NAIP5 sensor and nine NLRC4 helpers. While the ZAR1 resistosome is homogeneous, we cannot exclude the possibility that other plant NLRs can form heterogeneous protein complexes similar to NAIP5/NLRC4. CCs can form both homo- and hetero-oligomers comprising complex interaction networks (30, 86, 139), which would support the idea of heterogeneous resistosomes (Figure 2c). While CNLs often form hetero-oligomers (139), reports of TNL hetero-oligomers are relatively rare and are currently limited to paired NLRs (61, 74). It would be exciting to see if the composition of the paired NLR resistosomes is novel and has both sensors and helpers in equal proportions (Figure 3c).

The similarity between plant and mammalian cell death execution goes beyond structural similarities of resistosome to inflammasome. Recent reports of NADase activity of bacterial, mammalian and plant TIR domains (47, 48, 59, 131) suggest that activation of cell death can be mediated by conserved secondary messengers. Structural analyses revealed a conserved substrate binding site in the TIR domain of the mammalian executor of neuronal cell death SARM1 and plant TIRs (59). In line with the suggested common function of the TIR proteins across kingdoms, expression of the TIR domain from SARM1 in Nicotiana benthamiana produced a cell death response that was visually indistinguishable from HR (59, 131). The major difference between SARM1 and TNLs was revealed by their genetic requirements. Unlike plant TIRs, SARM1 was capable of inducing plant cell death independent of EDS1 and NRG1 (59, 131). The elucidation of the precise signaling cascade that culminates in cell death response can reveal further similarities across kingdoms as well as elucidate the plant-specific roles of EDS1 and NRG1.

Finally, does the solved resistosome structure represent its functionally active state? Research on mammalian NLRs suggests that inflammasome formation can be followed by proteasome-mediated release of the active caspase domain (104). Since it has been widely demonstrated in plants that N-terminal truncations of NLRs to either TIR domains or CCs alone are sufficient to induce cell death (16, 30, 52, 69, 122, 139), it is possible that the active resistosome is modified further in the plant cell before the induction of HR.
Exploiting effector activity

(a) Exploiting effector activities to engineer novel recognition. Prerequisites: a guardee and its corresponding NLR. The introduction of a known effector substrate site into the guardee or the NLR could yield durable resistance when deploying indispensable effectors. (b) Introducing new plant domains in NLR-ID. Prerequisite: an NLR-ID with relatively recent integration. IDs can be exchanged for plant protein domains that are known effector targets. (c) Harvesting natural diversity (indicated by different colors) to identify and subsequently engineer effector binding residues. Prerequisite: availability of information about sequence diversity over a recent period on direct binder NLRs that deploy variable LRR regions for the recognition of effectors. We hypothesize that residues involved in effector binding are more polymorphic and can therefore be identified through natural variation. These residues can be engineered to modify NLR binding specificity to confer altered or even novel recognition capacities.
What can we learn from common patterns observed in NLR evolution? One of the main practical ramifications of the observation that there are multiple ways to evolve the same function is that new functions can also be engineered in multiple ways. Engineering of these functions can be guided by learning the common principles of NLR biology.

4. CURRENT PROGRESS IN ENGINEERING NLR ALLELES

4.1. NLR Chimeras Define Effector Binding Regions and Allow Transfer of Recognition Between Allelic NLRs

Chimeras of flax NLRs were the first successful attempts in changing effector recognition of NLRs (46). This study analyzed 13 alleles of the flax rust resistance gene \( L \) (\( L_1 \)–\( L_{11} \) and \( LH \)) to identify variable regions (46). They subsequently created chimeras of \( L_2, L_6, \) and \( L_{10} \); transformed them into flax; and screened for changes in susceptibility to flax rust infection (46). The region exchanged between the NLRs contained 880 amino acids and included the entire LRR region. Transgenic flax plants expressing those chimeras showed resistance and susceptibility phenotypes correlating with the LRR included; in other words, the gene \( L_6-L_2 \), containing the \( L_6 \) TIR-NB-ARC combined with the \( L_2 \) LRR, displayed the resistance phenotype of the \( L_2 \) allele (46). This finding demonstrated that the LRR region can be sufficient to mediate recognition specificity of NLRs that directly interact with the effector they recognize. Since the first LRR swaps in \( L \), recognition specificity in several other direct binders has been mapped to the LRR region by either allelic swaps (114, 117) or in vivo interactions (26, 34, 44, 69, 102).

4.2. Introduction of New Effector Substrate Sequences into Plant Proteins

NLRs can directly bind effectors or recognize effector function by monitoring a host protein (i.e., a guardee) (Figure 1b). Since our understanding of effector functions is continuously improving, exploiting effector activities to engineer resistance seems like an elegant solution. Kim et al. (68) have used this strategy to engineer the guardee of RPS5 in order to expand recognition to several new effector modifications. RPS5 recognizes the action of the bacterial effector AvrPphB, which acts as a protease on PBS1 and related receptor-like cytoplasmic kinases (RLCKs). Upon cleavage by AvrPphB, these RLCKs undergo a conformational change that is recognized by RPS5 and leads to the activation of immune responses (43). The exchange of the proteolytic site in PBS1 with that of another effector, AvrRpt2, drastically improved resistance to \( P. syringae \) when expressed in otherwise susceptible \( Arabidopsis \) genotypes (68). The same strategy has also been used to insert the proteolytic site for the viral protease Nla of \( Tobacco etch virus \) into PBS1 (68). Although this approach led to cleavage of PBS1 through Nla and to HR in \( N. benthamiana \), it led to only partial resistance to the virus (68). This finding suggests that rates of substrate cleavage and strength of signaling need to be sufficient to achieve a robust response and may have different requirements for different effectors. Therefore, in some cases this approach might require additional optimization and protein engineering to yield resistance.

4.3. Introduction of Amino Acid Changes to Modify Effector Recognition Specificity

The discovery that the LRR can serve as an effector binding site and mediate NLR recognition specificity has prompted a number of attempts to randomly mutagenize LRRs to generate novel disease resistance. The NLR Rx provides resistance to \( Potato virus x \) strains carrying a variant
of its coat protein (CP) with the residues T121 and K127 (CP-TK). CPs with a lysine and an arginine in these respective positions (CP-KR) evade recognition by Rx (14). In one study (49), Rx was randomly mutated by means of error-prone polymerase chain reaction, and thousands of variants were screened for gain of recognition of CP-KR. This approach yielded several mutant Rx proteins that provided resistance against strains of *Potato virus x* carrying CP-KR and CP-TK, as well as against another related virus, *Poplar mosaic virus* (49). While this approach expanded the recognition of Rx, it did not yield novel resistance.

Two studies published in 2014 expanded the recognition capacities of the NLR R3a from the wild potato *Solanum demissum* by means of random mutagenesis, gene shuffling, and site-directed mutagenesis (33, 109). The wild-type version of R3a recognizes the *Phytophthora infestans* effector Avr3aKI but not the allelic variant Avr3aEM, which has become a prevalent allele in modern *Phytophthora* species, most likely due to positive selection mediated by evasion of R3a-mediated resistance (5, 20, 21, 130, 146). Segretin et al. (109) identified eight single amino acid substitutions that were able to trigger HR in *N. benthamiana* in response to Avr3aEM. Six of these mutations lie within the LRR region, one in the NB-ARC, and one in the CC. One of these mutations (K920E) was also identified in a second study (33). After several rounds of artificial evolution, Chapman et al. (33) identified several mutants that showed a stronger cell death response than the reference. However, these engineered NLRs, termed R3a+ and R3a*, were not able to provide enhanced resistance to *P. infestans* strains carrying Avr3aEM.

In a follow-up study, the mutations identified by Segretin and colleagues were transferred to the R3a ortholog I2 from tomato to see whether gain-of-function mutations are transferable between orthologs. I2 confers resistance to *Fusarium oxysporum* f. sp. *lycopersici* through the recognition of the Avr2 effector and has a high sequence similarity to R3a (98, 116). Two of the residues in the R3a+ alleles (I141F and N336Y) are conserved in I2 (60) and were subsequently mutated in I2 in order to expand its recognition capacities (53). However, transfer of these amino acid changes onto I2 led to either loss of recognition or autoactivity (53). To test the hypothesis that these two sites within the homologous NLRs could be hot spots for NLR sensitization, Giannakopoulou et al. (53) mutated both residues to all possible amino acid substitutions. This led to the identification of I2I141N, which can recognize both Avr3aKI and Avr3aEM, as well as two new Avr2 variants (53). Expression of these NLRs in *N. benthamiana* led to partial resistance to *P. infestans* carrying either variant of Avr3a (53). To date, therefore, attempts to randomly mutagenize NLRs have not led to the generation of novel disease resistance.

### 5. PROSPECTS AND CHALLENGES IN THE ENGINEERING OF NLRs

#### 5.1. Engineering Guardees for NLR Activation upon a Wide Range of Effector Activities

Exploiting effector activities to engineer novel disease resistance is a promising strategy, as pathogen effectors often evolve to evade NLR binding while retaining their activity. There are several ways in which we can exploit effector activities to the pathogen’s detriment. These include the strategy employed by Kim et al. (68) to transfer known effector-targeted proteolytic sites onto a plant protein that is guarded by a plant NLR. We might be able to engineer a guardee whose modification by several effectors can lead to NLR activation (Figure 3a). How many different types of effector activities can be recognized by a single guard NLR? The case of RIN4, which is guarded by several NLRs, including RPS2 and RPM1, and is targeted by several effectors with different enzymatic activities suggests that indirect recognition can indeed recognize only one modification, as RPM1 is activated only after phosphorylation of RIN4, while RPS2 recognizes...
its proteolytic cleavage (22). The fact that RIN4 is guarded by several NLRs makes it an attractive candidate for the introduction of more effector target sites. The study of effector enzymatic activities will continue to expand the list of potential targets for this approach (22). By choosing effectors with activities that are crucial for pathogen invasion, we can try to create durable resistance. This approach requires comprehensive knowledge of effector activities. Understanding natural effector targets and their activities is therefore crucial for this engineering approach and will greatly enhance our ability to create novel recognition capabilities.

5.2. Engineering NLRs with New Integrated Domains

In addition to modifying guards, we may be able to engineer NLR clades with IDs. Ideally, we could use existing NLR-IDs as platforms to create fusions with targets identified in effector interactome studies (Figure 3b) (101, 136). In practice, the NLR-IDs we observe today have domains that have coevolved for millions of years since the original fusion event. Therefore, this engineering approach will require careful dissection of coevolved regions or identification of a generalist NLR that can still accept new variable domain fusions, such as NLRs from MIC1, identified in Poaceae (10, 121). It will also require a better understanding of NLR-ID activation mechanisms. How does modification of the ID by the effector alter the conformation of the NLR-ID? What is the role of the helper NLR in the pair? Contrasting models have been proposed for different NLR pairs; one suggests that the helper suppresses autoactivity of the sensor NLR, whereas another suggests that the helper mediates initiation of immune signaling (61, 88). Improving our understanding of paired NLRs will facilitate the fine-tuning of immune activation of new NLR-ID fusions.

In the case of the allelic series of Pik genes in rice, the integrated heavy-metal-associated (HMA) domain is the most polymorphic region of the NLR (36), strongly suggesting that it is the effector binding interface. This idea was confirmed by crystallization of the Pik allele with its corresponding effector (88). A logical next step in utilizing effector binding to an ID is to exchange IDs for other known host domains targeted by effectors (Figure 3b). Such an exchange would enable the creation of recognition of effectors whose function remains elusive but whose targets have been identified through their introduction into existing NLR-ID scaffolds (Figure 3b). In addition to creating new fusion proteins, protein engineering of existing IDs to alter binding strength between the NLR and effector could be used to improve existing recognition. To this end, knowledge gained from crystal structures of NLRs in association with several allelic effectors, as performed by De La Concepcion et al. (40), has improved our understanding of binding interfaces and which amino acids to alter to increase binding affinity. This information will help guide engineering efforts to extend beyond the random mutagenesis approaches used in the past.

5.3. Exploiting NLRs with Naturally Diverse LRR Binding Specificities

The simplest mode of effector–NLR interaction is direct binding mediated through the variable LRR region, as it involves the fewest genetic components. As a result, direct binder NLRs are potentially the easiest class of NLRs to engineer (Figure 3c). However, we currently lack structural and biochemical data for these interactions. Key unanswered questions include the identity of residues within the LRR that determine binding specificity, the typical affinities required for productive interactions, and the structural rearrangements that lead to immune signaling. To sidestep this limitation, we could build upon research with allelic series in the flax L gene, discussed above (46), and look to natural allelic diversity and structural modeling to consistently identify binding pockets within the LRRs. By sampling the intraspecific natural diversity of NLRs, we should aim...
to determine sets of 10–20 residues within an LRR that are the most variable and, therefore, likely to be involved in effector binding. Such sets of residues will allow targeted engineering efforts to either improve existing binding specificities or derive new ones.

6. CONCLUSIONS AND FUTURE DIRECTIONS
Understanding the evolution of NLRs is crucial in order to determine their functionality, understand the environmental pressures influencing the plant immune receptor repertoire, and engineer new recognition specificities. It is now clear that we cannot apply knowledge of one NLR to another. The phylogenetic placement of NLRs, however, may indicate whether they are more likely to be helpers (e.g., RNLs) or sensors and what type of helpers they depend on (e.g., NRC-dependent clade of sensors). In addition, IDs not only highlight plant proteins that are potentially targeted by pathogens, and therefore monitored by the plant immune system, but also offer unprecedented potential for effector-guided engineering approaches. Nevertheless, it is important to remember that even among genes in one clade there can be functional differences, as is the case for NRG1 performance in Brassicaceae and Solanaceae (25, 71, 102). As sequence databases grow, we will be able to ask increasingly ambitious questions about NLR evolution and find common patterns that were previously concealed by noise.

Engineering of NLRs has proven difficult, and successful reports remain limited. On the basis of our current knowledge of NLRs, there are three main strategies to create novel disease resistance. The recent availability of sequence diversity of NLRs within populations provides the raw material for predictions on effector binding surfaces. For NLR-IDs, it is already apparent that IDs could potentially serve as effector target platforms. We now require an understanding of the timing of fusions and subsequent coevolution with canonical NLR domains. This information will inform the choice of a scaffold NLR that could be used for the exchange of IDs. For instance, NLR-IDs from the evolutionary youngest clade, MIC1, likely possess the greatest potential for successful engineering approaches. In the near future, lessons learned from NLR evolution and advances in synthetic and structural biology will enable us to create highly coveted designer NLRs.

SUMMARY POINTS
1. Subfunctionalization of nucleotide-binding leucine-rich repeat receptors (NLRs) into sensors and helpers arose independently multiple times across different species.
2. Helpers that evolved from NLR pairs are evolutionarily younger than coiled-coil domain NLR (RNL) helpers.
3. Subfunctionalization in sensor and helper signaling components of NLRs is reminiscent of the receptor and coreceptor relationship of pattern recognition receptors and might be a general evolutionary theme.
4. While predicting the mode of action of NLRs from sequence alone is challenging, phylogenetic placement in clades can help predict their genetic requirements.
5. Engineering of NLRs has been difficult, with only moderate success thus far.
6. Understanding the evolution of NLRs can inform engineering efforts and lead to designer NLRs with novel recognition specificity.
**FUTURE ISSUES**

1. Why do plants lose certain NLR clades but expand others?
2. Are evolutionary patterns conserved between intraspecific and interspecific scales?
3. How do human and pathogen evolutionary pressures affect immune gene evolution? Is this consistent across systems and between domesticated and wild systems?
4. On what timescale and by what means is the regulation of NLRs evolving in response to varying fitness costs with changing environments? Is this occurring at a faster or slower rate than sequence evolution within the NLR?
5. Why do NLRs subfunctionalize into helpers and sensors? How does this relationship work, and how is this complex activated by an effector?
6. What is the functional link between ENHANCED DISEASE SUSCEPTIBILITY 1 (EDS1), PHYTOALEXIN DEFICIENT 4 (PAD4), and RNL helpers? What are the exact molecular events leading to the hypersensitive response (HR), and what is its function in plant immunity?
7. What is the function of integrated domains (IDs)? Are they merely pathogen bait, or do they also fulfill a function by, for instance, targeting the NLR to the correct subcellular compartment?
8. Which NLR-IDs can serve as platforms for new fusion proteins? Which guardees can serve as platforms for effector target domains?

**DISCLOSURE STATEMENT**

The authors are not aware of any affiliations, memberships, funding, or financial holdings that might be perceived as affecting the objectivity of this review.

**ACKNOWLEDGMENTS**

We thank Dr. Paul Bailey for generously providing the NLR tree included as a figure in this review. We thank Erin L. Baggs and Dr. Daniil Prigozhin for insightful discussions of the concepts presented in this review and for critical reading of the manuscript. We also thank Pierre Joubert for providing his comments on the review.

**LITERATURE CITED**


49. Farnham G, Baulcombe DC. 2006. Artificial evolution extends the spectrum of viruses that are targeted by a disease-resistance gene from potato. *PNAS* 103:18828–33

---

68. Introduces an exogenous protease cleavage site into a guard cell to engineer resistance to a new pathogen.


---

113. Analyzes 22 angiosperms grouped in NLRs in three major clades and reconstructs their evolutionary history.

www.annualreviews.org • Evolution of Plant NLRs


141. Wu C-H, Kamoun S. 2019. Tomato Prf requires NLR helpers NRC2 and NRC3 to confer resistance against the bacterial speck pathogen *Pseudomonas syringae pv. tomato*. bioRxiv 595744. [https://doi.org/10.1101/595744](https://doi.org/10.1101/595744)


139. Shows that CC domains from *Arabidopsis* NLRs form a network of homo- and heteromeric interactions.

149. Analyzes 70 plants, showing the rapid birth of new miRNAs from recently duplicated NLRs.