Title

The Diversification of Plant NBS-LRR Defense Genes Directs the Evolution of MicroRNAs That Target Them.

Journal

Molecular Biology and Evolution (MBE), 33(10)

Authors

Zhang, Yu

Xia, Rui

Kuang, Hanhui

et al.

Publication Date

2016-10-01

DOI

10.1093/molbev/msw154

Peer reviewed

The Diversification of Plant *NBS-LRR* Defense Genes Directs the Evolution of MicroRNAs That Target Them

Yu Zhang,[1,2] Rui Xia,[2] Hanhui Kuang,[1] and Blake C. Meyers*[,2,3]

[1]Key Laboratory of Horticulture Biology, Ministry of Education, and Department of Vegetable Crops, College of Horticulture and Forestry, Huazhong Agricultural University, Wuhan, People's Republic of China

[2]Donald Danforth Plant Science Center, St. Louis

[3]Division of Plant Sciences, University of Missouri – Columbia

***Corresponding author**: E-mail: bmeyers@danforthcenter.org.

Associate Editor: Brandon Gaut

Abstract

High expression of plant nucleotide binding site leucine-rich repeat (*NBS-LRR*) defense genes is often lethal to plant cells, a phenotype perhaps associated with fitness costs. Plants implement several mechanisms to control the transcript level of *NBS-LRR* defense genes. As negative transcriptional regulators, diverse miRNAs target *NBS-LRRs* in eudicots and gymnosperms. To understand the evolutionary benefits of this miRNA-*NBS-LRR* regulatory system, we investigated the *NBS-LRRs* of 70 land plants, coupling this analysis with extensive small RNA data. A tight association between the diversity of *NBS-LRRs* and miRNAs was found. The miRNAs typically target highly duplicated *NBS-LRRs*. In comparison, families of heterogeneous *NBS-LRRs* were rarely targeted by miRNAs in Poaceae and Brassicaceae genomes. We observed that duplicated *NBS-LRRs* from different gene families periodically gave birth to new miRNAs. Most of these newly emerged miRNAs target the same conserved, encoded protein motif of NBS-LRRs, consistent with a model of convergent evolution for these miRNAs. By assessing the interactions between miRNAs and *NBS-LRRs*, we found nucleotide diversity in the wobble position of the codons in the target site drives the diversification of miRNAs. Taken together, we propose a co-evolutionary model of plant *NBS-LRRs* and miRNAs hypothesizing how plants balance the benefits and costs of *NBS-LRR* defense genes.

Key words: **NBS-LRR, disease resistance gene, plant, microRNA.**

Article

Introduction

Plants rely on multiple components of defense against microbial pathogens. Despite basal structural and chemical barriers, plant disease resistance is determined by two major classes of immune receptors that recognize appropriate ligands to activate defense (Jones and Dangl 2006). One class functions in basal defense, comprising pattern recognition receptors at the cell surface of plants, and recognizing conserved pathogen-associated molecular patterns (Monaghan and Zipfel 2012). A second class functions in elicitor-mediated defenses and comprises receptors that are typically cytoplasmic proteins containing nucleotide-binding site and leucine-rich repeat domains (NBS-LRRs), and detect polymorphic, strain-specific pathogen effectors (Eitas and Dangl 2010). *NBS-LRR* genes encompass the majority of resistance genes (*R*-genes) cloned thus far in plants. Plant NBS-LRR proteins belong to the STAND (signal-transduction ATPases with numerous domains) P-loop ATPases of the AAA+ superfamily (Bonardi et al. 2011). The central nucleotide-binding domain executes the function of a molecular switch which controls the ATP/ADP-bound state mediating downstream signaling (Takken and Tameling 2009). The N-terminal coiled-coil (CC) or Toll/Interleukin-1 receptor (TIR) domains are used as signaling hubs which can associate with either cellular targets of effector action or with downstream signaling components (Mukhtar et al. 2011). The C-terminal LRR domains of NBS-LRR proteins are variable in length and form a series of β-sheets with solvent-exposed residues which are believed to interact with specific ligands (Mondragon-Palomino et al. 2002; Maekawa et al. 2011).

Plant *NBS-LRRs* are polymorphic at the population level, varying both in the sequence composition of orthologs and in the overall number of paralogs. The sum of genes encoding *NBS-LRRs* in a host population defines the repertoire for the detection of polymorphic pathogen effectors (Kuang et al. 2004). *NBS-LRR* genes display two different patterns of evolution: type I genes are often represented by multiple paralogs in a genome and evolve rapidly with frequent gene conversions, while type II genes often have fewer paralogs and evolve slowly, with rare gene conversion events (Kuang et al. 2004). Type II genes are typically highly conserved and their variation may be limited to presence/absence polymorphisms, when compared across different individuals in a population. Most *NBS-LRRs* are organized in genomic clusters (Kuang et al.

2004; Luo et al. 2012). The number of *NBS-LRR* genes in different plant genomes varies substantially, from <100 to >1,000 (Yue et al. 2012; Xia et al. 2015). Some lineages of *NBS-LRRs* (i.e., *TIR–NBS-LRRs* and *NRG1* homologs) which are highly conserved in dicots are lost in monocots (Tarr and Alexander 2009; Collier et al. 2011). In general, the number of *NBS-LRR* genes is correlated with total number of genes in the genome (Wang et al. 2011). However, there are still numerous exceptions such as the quite low copy number of *NBS-LRRs* in the papaya, cucumber, and watermelon genomes (Lin et al. 2013; González et al. 2015). As NBS-LRRs can have fitness costs (Tian et al. 2003) and high expression of *NBS-LRRs* are often lethal to plant cells (Stokes et al. 2002), one hypothesis is that a high copy number of *NBS-LRRs* might be disadvantageous, restricting the number of active *NBS-LRRs* in a plant genome (Lin et al. 2013).

There is increasing evidence that small RNAs are involved in regulating plant immunity (Fei et al. 2016). Plant small RNAs are ~21–24 nucleotide (nt) regulators of gene expression, functioning at the transcriptional or post-transcriptional level in plants. It is well known that antiviral defense involves exogenous virus-derived small RNAs (reviewed in Zhu and Guo 2012). Recent work has revealed that host endogenous small RNAs (microRNAs and siRNAs) play roles in counteracting pathogens as well. For example, many microRNA families target *NBS-LRRs* (He et al. 2008; Zhai et al. 2011; Li et al. 2012; Shivaprasad et al. 2012; Xia et al. 2015). MicroRNAs (miRNAs) are generated from single-strand hairpin RNAs which are precisely processed and generate specific functional small RNAs—a duplex that includes the mature miRNA (Axtell 2013). Plant miRNAs, typically 21-nt in length, guide Argonaute proteins in a homology-dependent manner, predominantly to slice target mRNAs. Both miRNA and the accessibility of its target site within an mRNA affect the function (Kertesz et al. 2007; Shao et al. 2014); although an mRNA may have sequence homology to a miRNA, the sequence flanking the target site can play an important role in allowing or restricting access by the miRNA (Fei et al. 2015). MicroRNAs that are 22-nt in length, a somewhat aberrant size mostly resulting from precursors containing an asymmetrical bulge, can trigger the generation of secondary siRNAs from their target mRNAs, produced in a phased pattern (Chen et al. 2010; Cuperus et al. 2010). As a consequence of this pattern, these secondary siRNAs are known as phasiRNAs. In many plant genomes, large numbers of phasiRNAs are produced from *NBS-LRRs*, the result of 22-nt miRNAs that target conserved, encoded motifs (Zhai et al. 2011; Li et al. 2012; Shivaprasad et al. 2012; Fei et al. 2013). Yet the functional importance of small RNAs in *NBS-LRR* gene regulatory networks is still not fully understood. In an attempt to elucidate the benefits of this regulatory system, several recent articles investigate genomic diversity and correlations between miRNAs and *NBS-LRRs* (González et al. 2015; Zhao et al. 2015). Those analyses focused on a subset of genomes or species. Comprehensive analyses are needed to better understand how the miRNAs targeting *NBS-LRRs* originate and are maintained.

There are three main models for the emergence of new miRNAs in plant genomes. The first model is that a new microRNA is generated from the inverted duplication of a target gene sequence (Allen et al. 2004). The inverted duplication leads to the generation of a perfect complementary stem-loop, which is a substrate for processing by DICER-LIKE 4 (DCL4), yielding a group of 21-nt siRNAs that could regulate homologous genes such as the one from which the sequence was derived. Mutations subsequently accumulating in the stem-loop may disrupt the perfect complementarity, shifting the processing Dicer from DCL4 to DCL1 and simultaneously diminishing the number of small RNA duplexes produced, yielding a new miRNA. In this model, the amplification of a gene family is perhaps more likely to generate a miRNA that targets genes within this gene family. Our recent in spruce study shows that the precursor sequence of a copy of *MIR482* has significant similarity to *NBS-LRR* genes, which indicates that miRNAs targeting *NBS-LRRs* could emerge via this first model (Xia et al. 2015). The second model for the emergence of new miRNAs is that they arise from a serendipitously self-complementarity DNA sequence which by chance is transcribed and is Dicer-processed, yielding a small RNA that interacts with a homologous target; this small RNA–target interaction is fixed and refined through the co-evolution of miRNA sequence and target genes (Felippes et al. 2008). In the third model, a preexisting miRNA precursor evolves a new target specificity through co-evolution or perhaps punctuated jumps in sequence diversity, ultimately yielding a new mature miRNA that may bear little resemblance to its progenitor (Xia et al. 2013). In the context of these three models, the origins of many miRNAs now known to target *NBS-LRR* genes have not been fully assessed.

The link between *NBS-LRRs* and their regulation by small RNAs can be traced back to the emergence of gymnosperms, which was perhaps >100 million years after the origin of *NBS-LRR* genes in early land plants like mosses and spike mosses (Yue et al. 2012; Xia et al. 2015). At least eight families of miRNAs have now been described that are known to target *NBS-LRRs* (Fei et al. 2016). A common attribute among these families is that they target conserved regions, allowing one miRNA to target multiple lineages of *NB-LRRs*; for example, members of the miR482/2118 family target the encoded P-loop region. The presence of these miRNAs from gymnosperms to dicots indicates that these miRNAs originated prior to the emergence of angiosperms (Zhai et al. 2011; Shivaprasad et al. 2012). Other families of miRNAs targeting *NBS-LRRs* are younger, as evidenced by their presence only in specific plant lineages (He et al. 2008; Zhai et al. 2011; Li et al. 2012; Shivaprasad et al. 2012; Liu et al. 2014; Ouyang et al. 2014; Xia et al. 2015). To trace the evolutionary history of *NBS-LRR* genes and the miRNAs that target them, we analyzed the genomes of ~70 land plants, coupling this analysis with extensive small RNA data, allowing us to characterize the origin of *NBS-LRR*-targeting miRNAs. We show that miRNAs were continuously generated over evolutionary time, with the mature miRNA

sequence shaped by interactions with target *NBS-LRR* genes. The results suggest a genomic model of a dynamic relationship between *NBS-LRR* genes and miRNAs, characterized by expansion in *NBS-LRR* copy number, leading to the periodic emergence of miRNAs which might then act to contain potential fitness costs of their *NBS-LRR* progenitors.

Results

NBS-Encoding Genes in 2 of 35 Plant Families Are Significantly Heterogeneous

We identified genes encoding NBS proteins (hereafter, "*NBS* genes") similar to NBS-LRR disease resistance proteins, using 75 sequenced genomes representing 35 families of plants, including algal genomes, using a Pfam homology search (supplementary table S1, Supplementary Material online); a total of 20,571 *NBS* genes were identified, found only in land plants and starting with mosses. These genes were clustered into paralog/ortholog families by orthoMCL (Li et al. 2003), yielding 311 families (supplementary tables S2 and S3, Supplementary Material online), and leaving 4,492 genes as singletons, that is, not included in any of the 311 families. The *NBS* genes were classified into eight different types according to their annotation (fig. 1). Per genome, we observed that the copy number of *NBS* genes varied dramatically, yet when comparing related species, the proportion of different types of *NBS* genes was relatively similar (fig. 1). Both the number of *NBS* families ($r^2 = 0.51$, $P = 3.96 \times 10^{-12}$) and average copy number of genes of each family ($r^2 = 0.55$, $P = 1.71 \times 10^{-13}$) correlated with the total number of *NBS* genes among different genomes (supplementary table S4, Supplementary Material online). Such correlations indicate that the variation in *NBS* copy number among plants might result from a combination of each genome maintaining a diverse set of *NBS* families plus extensive gene duplication within each *NBS* family. In general, numerous *CC-NBS-LRR*s (*CNL*s) were present in all genomes while the *TIR–NBS-LRR*s (*TNL*s) were entirely absent in some lineages (fig. 1). The loss of *TNL*s in monocots, a major event in plant *R*-gene (*Resistance* gene) evolution (Meyers et al. 1999) must have occurred early after the divergence of monocots and eudicots since the early seed plant Amborella still has *TNL*s. We also confirmed the independent loss of *TNL*s in the eudicot *Mimulus*, as previously described (Collier et al. 2011). Finally, we noted a wide variation in the proportion of *TNL*s (of all *NBS-LRR*s) across eudicot genomes, varying from 0.55% to 54.17% (fig. 1, supplementary table S4, Supplementary Material online).

We next assessed the diversity of *NBS* genes in each genome, speculating that there may be a relationship between *NBS* gene diversity and pathogen resistance. Our measurement of diversity was simply the total number of *NBS* families (the paralog/ortholog families from orthoMCL, mentioned above), a quantification of the heterogeneity of encoded NBS-containing proteins in each genome. Across the land plants, we observed an average of ~29 distinct *NBS* families encoded per genome (28.8 ± 1.51). The Poaceae stood out for their 'global' composition of *NBS* genes, as the total number of *NBS* families encoded per genome was significantly higher than other plant families (61.5 ± 4.18; *t*-test, $P < 1 \times 10^{-8}$) (fig. 1). Next, within each of a subset of genomes (noted in fig. 1), we analyzed the clustering patterns of *NBS* genes since the organization of *NBS* genes in clusters on chromosomes is a common feature of their organization (Michelmore and Meyers 1998). The heterogeneity in each cluster was also measured by calculating the number of *NBS* families within a cluster (defined in the Materials and Methods); a heterogeneous *NBS* cluster contained *NBS* genes from different families, while a homogeneous cluster contained *NBS* genes only from one family. Comparing among different plant families, the proportion of heterogeneous versus homogeneous clusters was quite similar (supplementary table S5, Supplementary Material online). We next examined *NBS* gene pairs, immediately adjacent in a genome, focusing specifically on those in a convergent (or facing, $\rightarrow\leftarrow$) and divergent (or opposite, $\leftarrow\rightarrow$) transcription orientation, because for such gene pairs the expression is often correlated (Krom and Ramakrishna 2008). The proportion of *NBS* gene pairs found in a divergent orientation was much higher than those in a convergent orientation (71.3% divergent, 273/383). Finally, we looked at the heterogeneity of the pairs; heterogeneous pairs of *NBS* genes are often co-transcribed and function together to confer resistance (Williams et al. 2014; Saucet et al. 2015). In two plant families, the Poaceae (13.67 ± 3.55) and the Brassicaceae (14.23 ± 2.66), the number of heterogeneous convergent or divergent pairs of *NBS* genes was significantly higher than those in other genomes (4.5 ± 0.74) (*t*-test, $P < 1 \times 10^{-3}$) (fig. 1). Therefore, most plant families were relatively similar in their genomic arrangement of *NBS* genes, yet the Brassicaceae and Poaceae both stood out for essentially 'local' arrangements (heterogeneous *NBS* gene pairs).

To further examine genomic heterogeneity in NBS-encoding genes, we analyzed in each species two aspects of *NBS* gene family variation: (1) the presence/absence of different families, and (2) variation in copy number within each family (supplementary table S6, Supplementary Material online). The *NBS* families were designated with numbers from 1 to 311 according to the rank order of the number of genes in each family; *R* genes of known function were assigned to these families using BLAST and one was used as a reference gene name for the matched family (supplementary table S7, Supplementary Material online). Intriguingly, *NBS* genes from the 30 largest families (most duplicated) comprised 59.9% of our dataset; the gene copy number in the other 281 families was substantially lower (fig. 2A). This indicates that the genes in these 30 families are highly duplicated. The level of duplication within each family was variable across species. We next evaluated the preservation of *NBS* families by determining the number of species containing genes from that family. We found that, generally, the preservation of *NBS* sequences was poor since only 38 families were found both in eudicots and monocots. More importantly, the size of *NBS* families was not related to their preservation. This was true at both ends of

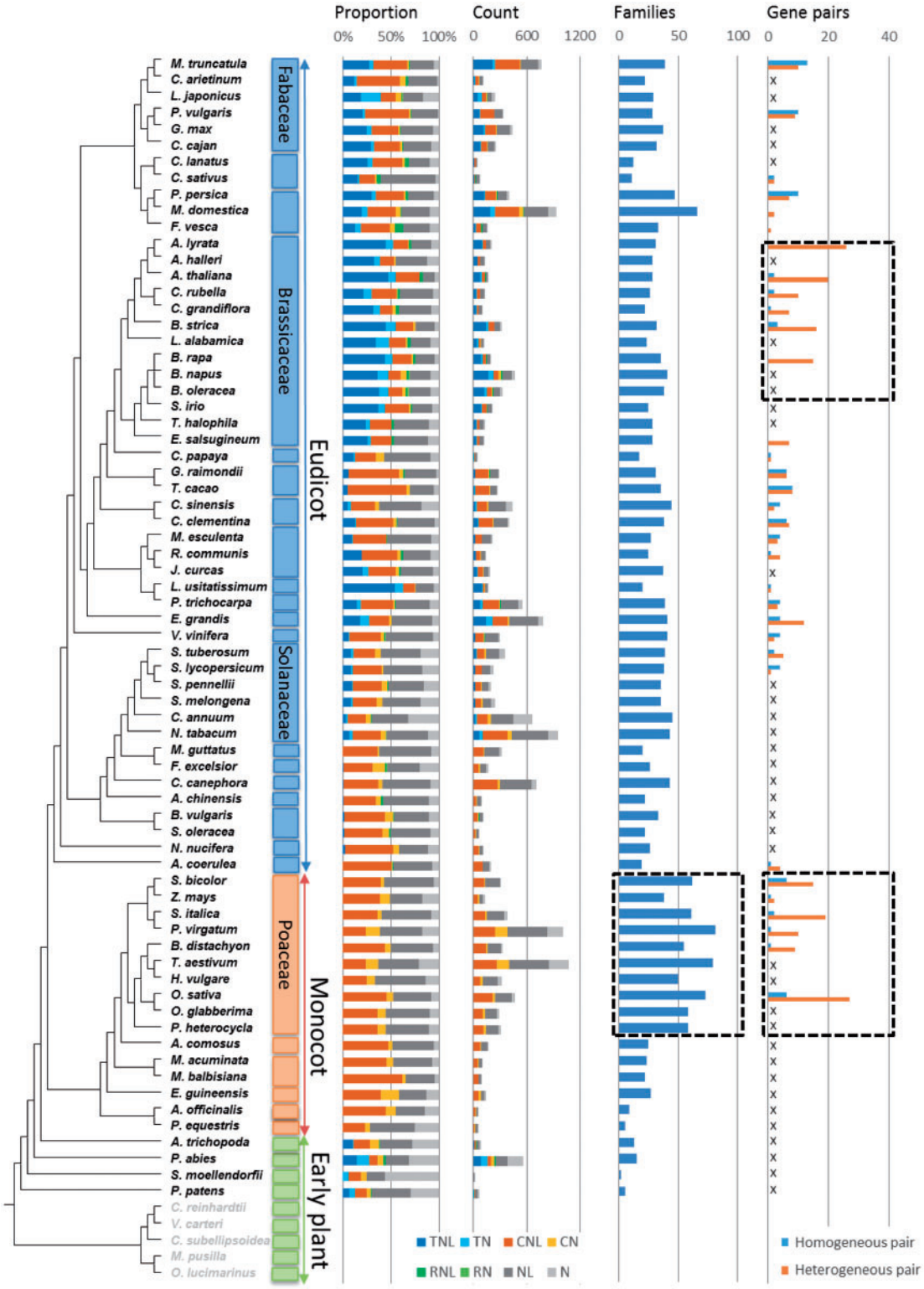
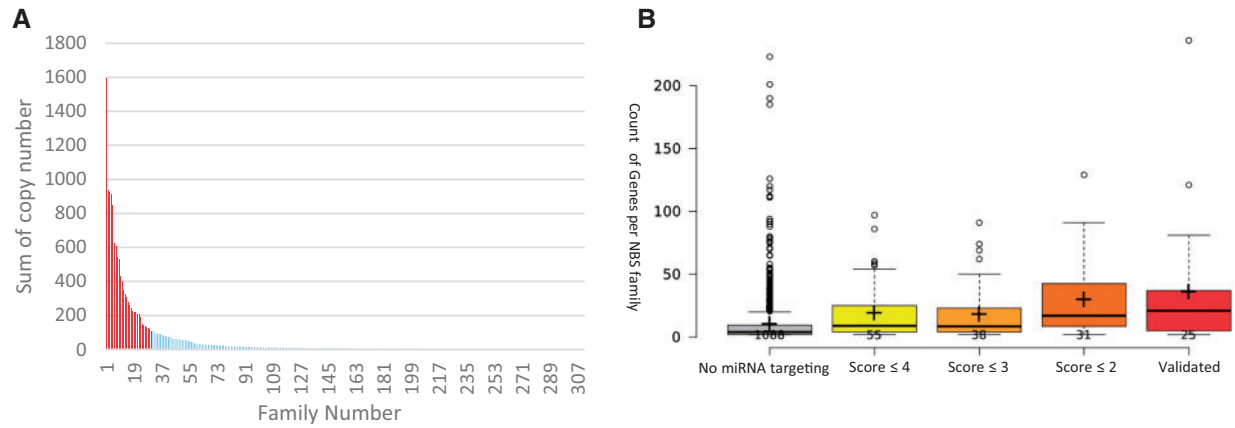


Fig. 1. Variation in copy number and composition of *NBS* genes in 75 plant genomes. The four bar charts, from left to right, represent for each genome the proportion of different types of *NBS* genes, the total number or count of *NBS* genes, the number of *NBS* families and convergent/divergent *NBS* pairs. The black boxes signify that total number of *NBS* families in the Poaceae and the number of heterogeneous pairs of *NBS* genes in the Poaceae and Brassicaceae is much higher than those in other genomes. In the chart for gene pairs, missing data are denoted with an "X" mark and is due to short genomic assemblies or the lack of a usable or complete gene annotation file.



C

	Family:	#1	#2	#3	#4	#5	#6	#7	#8	#9	#10	#11	#12	#13	#14	#15	#16	#17	#18	#19	#20	#21	#22	#23	#24	#25	#26	#27	#28	#29	#30
	Type:	TNL	CNL	CNL	CNL	CNL	CNL	TNL	CNL	CNL	TNL	CNL	CNL	CNL	CNL	CNL	CNL	CNL	CNL	CNL	TNL	CNL	CNL	CNL	RNL	RNL	CNL	CNL	CNL	TNL	TNL
	Reference:	*N*	*FOM-2*	*Pi-ta*	*RPM1*	*Dm3*	*RPS5*	*RPP5*	-	*RPP8*	-	*Pit*	-	-	*Prf*	-	-	-	*Hero*	*PIB*	*Rps4*	*RPP13*	-	*Bs2*	*NRG1*	*ADR1*	*Lr1*	*Pm3*	-	*RRS1*	-
M. truncatula		236	81	0	54	97	4	0	1	2	55	0	3	19	0	0	35	0	0	0	0	1	20	0	8	2	0	0	0	0	0
G. max		121	14	0	55	34	1	0	7	4	19	0	1	23	0	0	8	0	0	0	0	2	8	0	4	4	0	0	0	0	0
P. persica		91	22	0	38	12	5	0	0	25	40	0	0	9	0	19	6	0	0	0	0	4	3	0	10	1	0	0	0	0	0
M. domestica		129	58	0	44	69	7	0	0	14	90	0	0	20	0	60	29	0	0	0	0	1	23	0	16	5	0	0	0	0	4
A. lyrata		1	0	0	1	0	17	33	0	16	1	0	0	0	0	0	0	0	0	0	25	2	0	0	2	3	0	0	0	15	0
A. thaliana		2	0	0	1	0	21	36	0	16	1	0	0	0	0	0	0	0	0	0	11	3	0	0	2	4	0	0	0	9	0
B. rapa		5	0	0	0	0	25	40	0	16	1	0	0	0	0	0	0	0	0	0	15	1	0	0	5	4	0	0	0	10	0
C. papaya		7	1	0	1	3	0	0	0	3	2	0	0	3	0	4	3	0	0	0	0	0	0	0	0	0	0	0	0	0	0
G. raimondii		16	35	0	1	47	23	0	0	24	9	0	45	11	0	6	5	0	0	0	0	0	4	0	1	3	0	0	0	0	0
C. sinensis		69	4	0	6	57	60	0	0	1	0	0	11	9	0	50	22	0	0	0	2	5	9	0	2	1	0	0	0	0	0
P. trichocarpa		86	74	0	24	63	3	0	0	3	46	0	77	20	0	24	9	0	0	0	1	0	3	0	3	2	0	0	0	0	3
S. tubersosum		25	30	0	5	10	0	0	37	2	11	0	3	2	13	1	3	0	24	0	0	21	5	15	0	0	0	0	0	0	0
S. lycopersicum		13	7	0	3	4	0	0	30	2	6	0	3	1	12	1	0	0	24	0	0	11	1	5	1	1	0	0	0	0	0
S. bicolor		0	1	62	24	0	1	0	0	0	0	19	0	2	0	0	0	3	0	15	0	0	0	0	0	0	13	11	8	0	0
Z. mays		0	0	20	14	0	1	0	0	0	0	5	0	0	0	0	0	3	0	4	0	0	0	0	0	1	1	1	3	0	0
O. sativa		0	0	91	15	0	1	0	0	1	0	37	0	0	0	0	1	4	0	24	0	2	0	0	0	1	10	2	6	0	0
B. distachyon		0	0	82	38	1	2	0	0	0	0	31	0	5	0	0	0	18	0	8	0	0	0	0	0	1	12	13	6	0	0
A. trichopoda		0	0	0	9	0	1	0	0	3	1	0	0	0	0	0	0	0	0	0	0	1	0	0	0	1	0	0	0	0	4
P. abies		0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	31	0	0	0	0	69
P. patens		0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	23

Validated Score ≤ 2 Score ≤ 3 Score ≤ 4

Fig. 2. Highly duplicated *NBS* genes targeted by miRNAs. (*A*) Count of genes in each *NBS* family; the complete data are in supplementary table S6, Supplementary Material online. Families #1–#30 are labeled in red to indicate their high level of duplication (exceeding a threshold of 110 copies). (*B*) The correlation between miRNAs that target *NBS* genes and the average count of genes (per *NBS* family). The five categories are as follows: (red) those having at least one validated miRNA target (this is the strongest category); (brown, orange, yellow, respectively) those with at least one predicted miRNA targeting with the best alignment score $\leq$2, 3, 4; or (grey) no miRNA targeting (including poor miRNA–*NBS* interactions). (*C*) The 30 largest families of *NBS* genes (at top) measured across 20 representative plant genomes, as indicated at left. The number in each block indicates the count of *NBS* genes from a given family in a particular species. The colored blocks use the same colors as in panel B to denote the best alignment score between miRNAs and *NBS* target sequences; a lack of color indicates that no predicted target at or below the threshold score of 4. The references for each family are the BLAST hits of functional *R* genes (supplementary table S7, Supplementary Material online). Brassicaceae and Poaceae species in the left column were highlighted in green and brown.

the spectrum; for example, the two largest families (#1, *N*-gene, and #2, *Form-2*) are restricted in certain lineages. They are likely to encode immune receptors for which duplication might be beneficial by increasing resistance to more pathogens. In contrast, the two most-conserved *NBS* families (#25, ADR1 and #4, RPM1) were both relatively low copy in plant genomes, particularly in comparison to the families named *N*-gene and *Form-2*. In the Poaceae and Brassicaceae, the highly duplicated *N*-gene and *Form-2* NBS families were either absent or found in very low numbers (supplementary table S6, Supplementary Material online); coupled with the relatively high level of heterogeneity in the *NBS* genes, these are clearly unusual plant families.

Highly Duplicated *NBS* Genes Are Typically Targeted by miRNAs

Previous reports suggested that miRNAs may preferentially target duplicated *NBS-LRRs* (González et al. 2015; Vries et al. 2015), particularly those that are tandemly duplicated (Zhang et al. 2014). One of these reports proposed that plant genomes encoding a large number of *NBS-LRRs* are also those having *NBS-LRR*-targeting miRNAs in eudicots (González et al. 2015). Another report evaluated the relationships between miRNA targeting and functional diversity of the *NBS-LRR* repertoire in the Solanaceae (Vries et al. 2015). These analyses assessed limited numbers of miRNAs and species.

Therefore, we sought to build on this study by integrating data on the duplication of *NBS* gene families, to examine the larger number of plant genomes available today, and to include a more extensive set of miRNAs. To achieve this, of the original 70 land plants for which we examined the NBS-encoding genes, we focused on 20 representative land plants with sequenced genomes and good small RNA data (fig. 2). For the 20 genomes, we analyzed miRNAs from miRBase or prior publications (supplementary data S1, Supplementary Material online); the subset of miRNAs predicted to target *NBS* genes was identified by calculating the miRNA-target alignment score (using TargetFinder) and flanking sequence accessibility (using RNAup; see Materials and Methods for the criteria). Finally, 2,811 pairs of miRNA–*NBS* interactions were identified using a matching score cut-off ≤ 4, among which 1,645 pairs were regarded as poor since their target accessibility is >-20 kcal/mol (this cut-off was based on a prior study, Shao et al. 2014) (supplementary table S8, Supplementary Material online). We found a negative correlation between the matching score and flanking sequence accessibility ($r^2 = 0.14$, $P = 3.22 \times 10^{-96}$), which indicates that both components of the interaction (targeting and accessibility) have co-evolved. The cleavage of target sites was validated by PARE data or the presence of phasiRNAs (an outcome characteristic of 22-nt miRNA-guided target slicing) in *Arabidopsis*, *Medicago*, soybean, tomato, *Brachypodium* and spruce (supplementary table S8, supplementary data S2, Supplementary Material online). The complete set of NBS families was then classified into five categories based on the strength of target prediction and validation (fig. 2B). We plotted these categories compared with the count of gene copies for all NBS families; we observed a strong relationship between the number of *NBS* genes and miRNA targeting, as previously described (Vries et al. 2015), with the most robust miRNA–*NBS* interactions (i.e., those validated, previously or by us), having the highest average copy number per gene family (fig. 2B).

We individually investigated the miRNA–*NBS* interactions for each plant species, using the 30 largest *NBS* families (#1–30; fig. 2C). In general, we observed that within a family, for *NBS* genes highly amplified in a genome, there is a high probability that those genes are targeted by miRNAs. For example, genes in families #1, 2, 4, 5, 6, 8, 10, 12, 13, 15, 16 and others are widely targeted by miR482 (but not in the Brassicaceae and Poaceae). In fact, our conclusion is strongly influenced by miR482, a large and complex family of miRNAs (Shivaprasad et al. 2012; González et al. 2015) present in most of the genomes we examined. Excluding miR482, many highly duplicated *NBS*s are targeted by lineage-specific miRNAs. For example, the ADR1 homologs (family #25) are the most well-retained among different species, are present in only a few copies (one to five copies) in most plant genomes, and are rarely targeted by a miRNA. However, in spruce, there are 31 *ADR1* homologs; miR3693, a miRNA found thus far only in spruce, was predicted to target this set of genes (Xia et al. 2015). For the Poaceae and Brassicaceae, the two plant families described above as having unusual characteristics, represented the same case: for highly amplified *NBS* families, we typically observed an association with a lineage-specific miRNA that targets it. For example, in the Brassicaceae, although families #1–5 are duplicated inefficiently (i.e., present in low copy numbers) in *Arabidopsis* and *Brassica* and have no targeting miRNAs, families #6 and #7 which are highly duplicated yet are targeted by miR472 (#6), and miR825 & miR1885 (both target #7). In the Poaceae, there are at least two validated miRNAs targeting high-copy number *NBS-LRR*s: *Brachypodium*-specific miR5163 targets *NBS* genes from family #11, triggering secondary phasiRNAs, and the wheat-specific miR9863 targets family #3 (Liu et al. 2014). Possibly the Brassicaceae and Poaceae rarely have miRNAs targeting their *NBS-LRR* genes (Fei et al. 2013) because the large, duplicated gene families widely targeted by miR482 are poorly amplified in these two unusual plant families.

The Evolutionary Origins of miRNAs from *NBS* Genes Supports a Model of Convergent Evolution

We next investigated the models that describe how miRNAs might emerge over evolutionary time. As described above, there are three models: (1) from the inverted duplication of a target gene sequence, (2) from a serendipitously self-complementarity DNA sequence which by chance is transcribed and is Dicer-processed, (3) from divergence of a preexisting miRNA precursor. The correlation between miRNA targeting and highly duplicated homogeneous *NBS* genes suggests that *NBS–LRR*-targeting miRNAs might emerge via the inverted duplication of a target gene sequence. To test this hypothesis, the hairpin sequence of each miRNA was used to search via BLASTN the full dataset of *NBS* genes for matches, using an *e*-value cutoff of $<10^{-5}$. This identified miRNA precursors with extended similarity (beyond the mature miRNA sequence) to *NBS* genes, including *MIR472*, *MIR825*, *MIR1885*, *MIR6025*, *MIR1510*, *MIR2089*, *MIR5163*, and *MIR9863* (supplementary table S9, Supplementary Material online). We cross-checked the result by BLASTN against all genes in Genbank, and the top-matched genes were consistent (data not shown). An inspection of a subset of these miRNAs illustrates this match (fig. 3A). Intriguingly, these miRNA precursors matched genes in the same NBS family with their best target genes. For example, in *Arabidopsis lyrata*, miR472 and miR825 target genes from families #6 and #7, respectively; the best hit of the 5′ arm of aly-*MIR472* is *siri*|*NBS0045* and 3′ arm is *alyr*|*NBS0176* (*NBS* identifiers are listed in supplementary table S2, Supplementary Material online), which are also from family #6. Similarly, the best hits for the 5′ and 3′ arms of aly-*MIR825* were *brap*|*NBS0179* and *fves*|*NBS0079*, respectively—both from family #7. In addition, the inverted repeat elements around *NBS-LRR* loci from the apple, cotton, *Medicago* and tomato genomes were identified, to seek evolutionary intermediates in transition from inverted pairs of genes to miRNAs. We identified 15, 6, 16, and 7 *NBS* genes, respectively, containing at least one inverted repeat covering their exons. Among the inverted repeats, six were located at the junction of two convergent or divergent pairs of *NBS-LRR* genes. These inverted repeats can form good fold-back

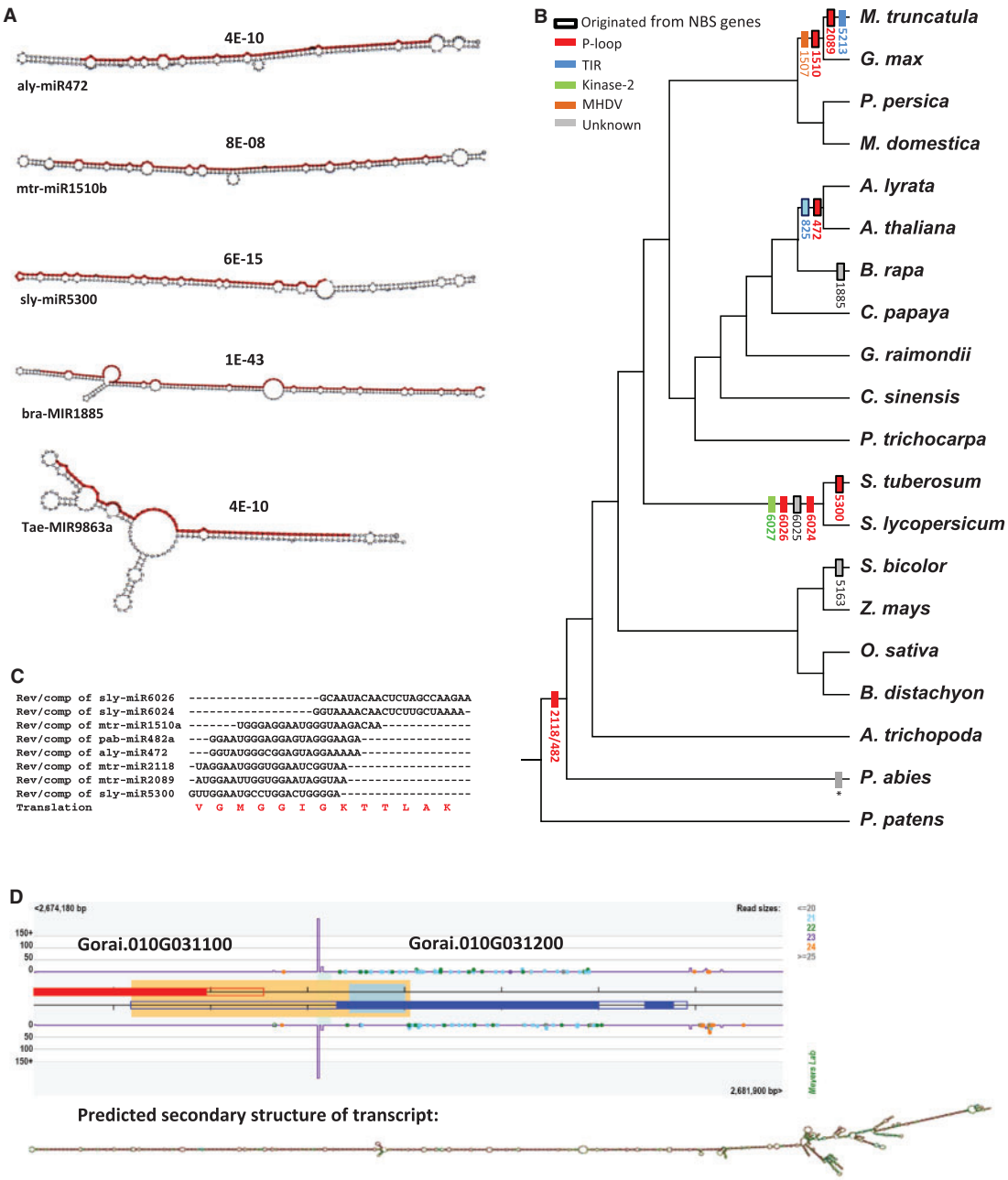


FIG. 3. Convergent evolution of miRNA-targeting *NBS-LRRs*. (*A*) Some miRNA precursors have extended similarity with *NBS* genes beyond miRNAs. The red highlighted sequence represents a miRNA precursor with a best BLAST hit to an *NBS* gene. The number above the hairpin is the *e*-value of the best BLAST hit. (*B*) MicroRNAs continuously emerged during the evolution of higher plants and target encoded, conserved motifs of NBS proteins. The boxes represent the birth of a miRNA during the differentiation of plant ancestors, as determined from the earliest lineage in which it was identified. Boxes with a black outline indicate miRNAs with evidence of origination from an inverted repeat of an *NBS* gene. Red, blue, orange, and green miRNAs target the encoded P-loop, TIR, Kinase-2, or MHDV motif of *NBS* genes, respectively. The asterisk for spruce indicates many diverse miRNAs target *NBS-LRRs* (Xia et al. 2015). All the red identifiers indicate separate origins of miRNAs that target the same P-loop-encoding region of *NBS-LRRs*. (*C*) Alignment of reverse-complement of the sequence of miRNAs targeting *NBS* genes, indicating their target sites around the encoded P-loop motif (the amino acids for this are shown at the bottom). (*D*) Above, a potential siRNA-producing locus formed by two inverted repeats of a homogeneous *NBS-LRR* gene pair. The red and blue bars indicate two annotated *NBS* genes as a convergent pair in the cotton genome. The orange box indicates a predicted inverted repeat. The small dots illustrate 21–24-nt small RNAs in the region (according to the legend in the upper right). Below, the MFE secondary structure of the inverted repeat was predicted by RNAfold.

structures, which could be both potential siRNA loci and recently-emerged precursors for miRNA genes (fig. 3D, supplementary data S3, Supplementary Material online). We interpreted this result to mean that these miRNAs originated from specific 'parental' *NBS* genes, and, due to sequence similarity, targeted their relatives.

The presence/absence of *NBS*-targeting miRNAs across a phylogeny of plants was examined to characterize the time

points at which miRNAs emerged. The miR482/2118 superfamily is exceptional, as it is clearly the most ancient of these miRNAs, probably generated during the emergence of seed plants (fig. 3B). While our analysis has constraints in terms of resolution, it appears that all other miRNAs evolved during the divergence of a specific plant family or species (fig. 3B). For example, miR472, miR825, and miR1885 are apparently specific to the Brassicaceae; miR1507, miR1510, miR2089, and miR5213 to the Fabaceae; miR6024, miR6025, miR6026, miR6027, and miR5300 to the Solanaceae. Spruce stands out for the large number of apparently lineage-specific miRNAs targeting *NBS-LRRs* (Xia et al. 2015). Across all examined plant species, conserved sequences within the *NBS* genes were the typical targets of these miRNAs, encoding motifs such as the P-loop, TIR, Kinase-2 and MHDV, consistent with prior observations from individual species (Fei et al. 2013). Among the encoded motifs, the P-loop was predominant, the target of miRNAs that include miR6024, miR6026, miR1510, miR2089, miR5300, miR472, miR482, and miR2118 (fig. 3C). This is quite remarkable because these miRNAs that share a common target site evolved at quite different times and in different lineages, as indicated in figure 3B. We observed that even miRNAs with closely related mature sequences could emerge over a broad evolutionary time period, in divergent species (fig. 3C). The best example is miR472, considered to be part of an ancient family that includes miR482 and miR2118 (Shivaprasad et al. 2012; González et al. 2015; Xia et al. 2015). However, miR472 clearly evolved recently since its backbone still has a high degree of similarity to *NBS* genes (supplementary table S9, Supplementary Material online), while other miR482/2118 superfamily members demonstrated no similarity to *NBS* genes (outside of the mature miRNA sequence). Though these miRNA families targeting *NBS-LRRs* are continuously generated over millions of years, the process by which they emerged is consistent (via inverted duplication of target sequences). That these miRNA families have distinct, lineage-specific origins yet in many cases target the same encoded motif (i.e., fig. 3C) is perhaps indicative of convergent evolution.

Amino Acid Diversity of the P-Loop Is a Major Force Driving the Evolution of the miRNA–Target Interactions

The relationship between the encoded P-loop of *NBS* genes and miRNAs is unique, with a complex interplay between conservation and the diversification of nucleotides and amino acids. We next assessed the relative influence of the protein versus miRNA targeting during evolution at this site. In other words, we asked how the interaction between a miRNA and its target site is maintained over 400 million years of evolution, asking in particular the direction or balance of the selective forces acting on the encoded protein sequence versus on the miRNA target site. The focus of our investigation was the most ancient interaction: that of miR482/2118 and its target, the encoded P-loop motif of *NBS* genes. The input data for this analysis was the same set of 20,571 *NBS* genes from the 70 plant genomes; for the P-loop motif encoded by each gene, we extracted the seven amino acids that form the core of the motif and are encoded by the target site of the miR482/2118 family. We observed a small number of major variants of the seven amino acids in the P-loop core, but with substantial variation in genic frequency of these variants among plant families (supplementary table S10, Supplementary Material online). In most eudicots, the encoded P-loop core in the *NBS* genes was frequently GMGGIGK or GMGGVGK, but in some Brassicaceae species, the most frequent P-loop core was one not common in other eudicots, GPAGIGK (supplementary table S10, Supplementary Material online). In monocots, the most frequent P-loop core was GFGGLGK or GMGGLGK (supplementary fig. S1, Supplementary Material online). The *NBS* genes with encoded P-loops of GMGG[V/I]G (i.e., the eudicot type) are preferentially targeted by miR482/2118; an alignment of many members of this family in eudicots, reverse complemented and translated, showed a very high degree of conservation (supplementary fig. S1, Supplementary Material online). This variation across plant families of the types of P-loop core sequences and in miRNA target preferences led us to ask which might have played a stronger role in driving selection.

To determine the directionality of evolution at P-loop core site, that is, whether selection is driven by either the protein or the miRNA, or is bidirectional and balanced, we first addressed the possibility of selection driven by the miRNA that could influence the nucleotide diversity at the core of the P-loop. If the miRNA predominates in evolution and thus constrains diversity at its target site, we might expect a reduced level of synonymous substitutions in the encoded P-loop (nucleotide changes that do not alter the encoded amino acid). We compared other species with Poaceae and Brassicaceae, knowing that they generally lack targeting of *NBS* genes by miR482/2118 (see above). For all families, we calculated the proportion of synonymous (P_s) and nonsynonymous (P_n) substitutions at the encoded P-loop core. There was no significant difference among plant families in the proportion of synonymous substitutions (supplementary fig. S2 and table S11, Supplementary Material online). Therefore, miRNAs may not strongly influence P-loop nucleotide diversity. However, while the proportion of non-synonymous substitutions did not vary substantially in most families, we observed a notably higher level (*t*-test, $P < 0.01$) at the encoded P-loop core of the Poaceae (0.2936 ± 0.0079) and the Brassicaceae (0.2726 ± 0.0069) relative to other plant families (0.2382 ± 0.0062) (supplementary fig. S2 and table S11, Supplementary Material online). This higher level of diversity at the P-loop core (i.e., non-synonymous substitutions) is perhaps a direct result of the relatively higher heterogeneity of *NBS* genes in these two plant families that we described above, rather than the absence of miRNA targeting. Another line of evidence suggesting that the miRNA targeting is not strongly driving evolution of the protein resulted from our analysis of the accessibility of the miR482/2118 target sites at the encoded P-loop core. We observed that the accessibility of this target site in *NBS* genes was frequently quite low (> -20 kcal/mol)

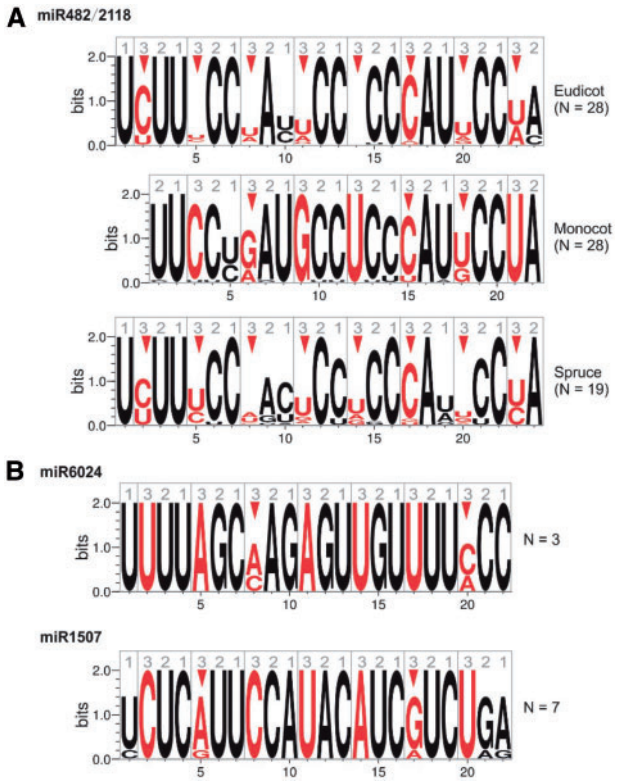


FIG. 4. Selection from target site to miRNA. (*A*) The diversity of miRNAs targeting *NBS* genes, measured in codon-based increments. The sequence logo presents the alignment of miRNA sequences in the miR482/2118 family. The "N = #" entries describe the number of sequences used in each alignment. The red nucleotides are in the third position of codons at the target site, with the numbers above in grey indicating these target sequence codon positions. The red arrowheads indicate the positions that both show diversity and match the third position of a codon in the encoded P-loop sequence. (*B*) Two example miRNAs that were selected because they are lineage-specific and presumably recently evolved, yet are distinct from the miR482/2118 family. With the exception of 5′ and 3′ nucleotides, diversity occurs in the third codon position.

(supplementary fig. S3, Supplementary Material online). This pattern is the same for other families of miRNAs targeting *NBS* genes (supplementary fig. S2, Supplementary Material online), indicative of relaxed selection of the target site by the miRNA, resulting in the frequent loss and gain of target accessibility. In conclusion, in eudicots, there might be no or low selection driven by the miRNA that could influence the nucleotide diversity at the core of the P-loop.

We next addressed the possibility of selection driven by the NBS protein, directing sequence diversification among nucleotides encoding the P-loop core more strongly than the selection pressure of miRNAs that target the same site. This analysis of miRNA sequences was performed relative to the codons in the encoded P-loop at the target sites; we hypothesized that if the miRNA diversity reflected codon degeneracy, then the miRNAs likely diversified to regulate a broad range of NBS-encoding genes. Alternatively, if there was little evidence of codon degeneracy in the miRNA alignments, this may indicate that target diversification was constrained by the interactions with the miRNAs. In total, 75 copies of miR482/2118 were aligned, divided into three subgroups: those from eudicots, in which these miRNAs are known to target only *NBS* genes; those from the Poaceae, in which these miRNAs largely target long, non-coding RNAs (lncRNAs) (Johnson et al. 2009; Zhai et al. 2015); and those from spruce, a gymnosperm in which these miRNAs target both *NBS* genes and lncRNAs (Xia et al. 2015). Within these alignments, we assessed the nucleotide diversity among the family members, relative to target sequence codons (fig. 4A). For the eudicot copies, across the entire alignment, almost all the diversity corresponded to the third position of the target site codons (fig. 4A). In contrast, in the alignment of the Poaceae miRNAs, a much weaker periodic pattern was observed for the third codon position in the targets (fig. 4A); this is notable because of the predominance of lncRNA targets in the Poaceae. The pattern in spruce was more consistent with eudicots. We also examined two recently-emerged miRNAs to assess how their diversity was shaped by their co-evolution with *NBS* targets. We examined miR1507 and miR6024, which are miRNAs found only in the Fabaceae and Solanaceae, respectively, that are distinct from the miR482/2118 family (Zhai et al. 2011; Li et al. 2012). Analyzed as above for miR482/2118 (i.e., relative to the target codons), while there were few positions that showed much diversity, the diversity that we observed was concentrated at third codon positions in the target sites (fig. 4B). We concluded that for the *NBS*–miRNA interactions of the eudicots, the high level of conservation in the miRNAs matching with the first two positions of each codon, and the diversity in the third position, are consistent with selection acting more strongly at the amino acid level (i.e., the P-loop), with miRNAs diversifying to target the broadest range of more-constrained, protein-encoding targets.

Discussion

The evolution of disease resistance has been a topic of interest in the eras of both classical and molecular genetics. Approximately 20 years ago, the first *NBS-LRR* gene was cloned (Whitham et al. 1994), with many functional *R* genes identified since that time. These genes were subsequently studied for their diversity and evolutionary rates. The first miRNA targeting *NBS-LRR* genes that was identified was miR472, reported in *Arabidopsis* (Lu et al. 2006); since 2006, numerous other miRNAs targeting *NBS-LRRs* have been described (Fei et al. 2016). In our study, we examined the connection between *NBS-LRRs* and the miRNAs that target them, working with 70 available land plant genomes. It is possible that the miRNA targets and gene count in these species might be affected by (1) the completeness and accuracy of the genome sequences, and (2) using miRBase as the main miRNA dataset resource; yet, our conclusions were based mainly on well-described gene families and validated miRNAs, supported by the large-scale genomic analysis. The data were consistent with the same primary points: both the miRNAs and *NBS* genes are highly dynamic genomic elements, as evidenced by the diversity of *NBS* genes, the

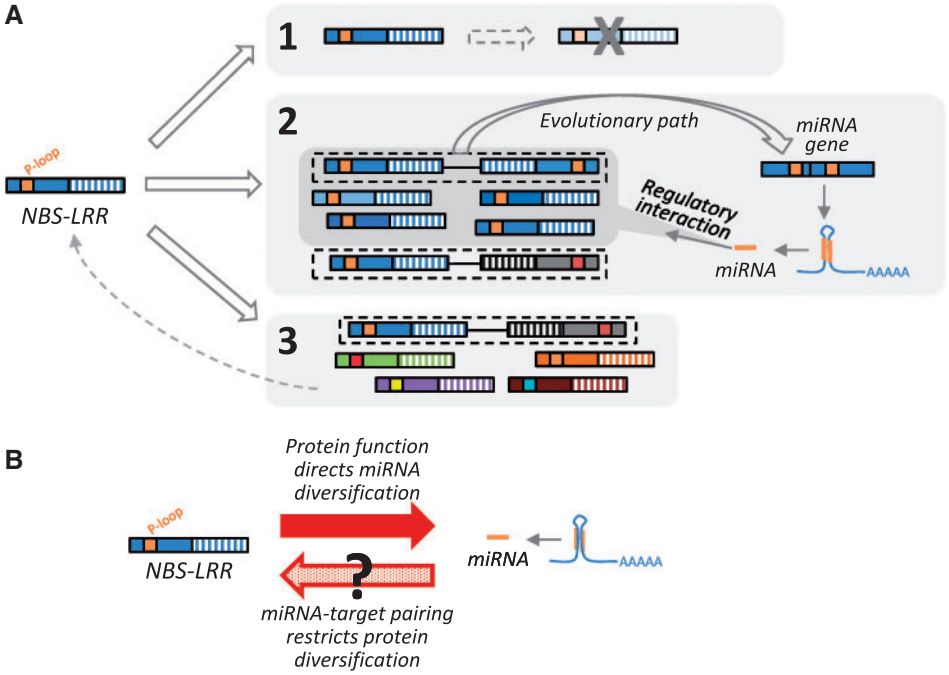


Fig. 5. Co-evolutionary model of plant NBS-LRR-encoding genes and miRNAs. (*A*) *NBS-LRR* genes may have evolved via three paths, related to their copy number and diversity. (1) Some *NBS* genes endure over evolutionary time as a single copy (or in low numbers), and their allelic diversity may be limited to presence/absence variation (the deleted copy denoted by the X-marked gene at the right). (2) Direct duplication of *NBS-LRR* genes can form complex clusters, and larger families of related genes at a genome scale. Highly duplicated *NBS-LRR* genes may have a higher probability to generate a new microRNA via localized genomic rearrangements; these miRNAs can then target and regulate the family of *NBS-LRR* genes from which they are derived. The dashed line indicates two adjacent *NBS-LRR* genes; a related pair in an inverted tandem duplication, with small genomic rearrangements, could yield a miRNA-like inverted repeat, whereas a heterogeneous pair (two distantly related *NBS-LRR* genes—indicated below) in the same formation are unlikely to yield a miRNA-like precursor, and thus may ultimately fail to evolve miRNA regulation. The encoded, conserved P-loop motif plays a central role in maintaining the relationship between the emergent miRNAs and their target *NBS-LRR* genes. (3) Some duplicated *NBS-LRR* genes may diversify extensively, emerging as distinct individual or families of *NBS-LRR* genes. These *NBS-LRR* genes may re-initiate the cycle of divergence, as indicated by the thin arrow back to the starting point. (B) We examined the relationship between *NBS-LRR* diversification and the miRNA–target interactions. We asked whether the dominant component of the relationship is (at left) the function of the protein, and therefore the constraints on gene diversification are imposed by the protein function, or by (at right) the miRNA that targets *NBS-LRR* genes, frequently at the P-loop encoding sequences. Our data indicate that protein function is a strong force (red arrow) that is directing diversification of miRNAs; it is still unclear how strongly miRNAs constrain diversification of the proteins encoded by their *NBS-LRR* gene targets.

emergence of new miRNAs that target these genes, and the demonstration that the forces of selection shaping these genes have been complex.

Expansion, Diversification, and Preservation of Plant *NBS-LRR*s over Deep Evolutionary Time

*NBS-LRR*s apparently originated with mosses, and diversified in the seed plants (Yue et al. 2012). Among different plant species, we observed that different *NBS-LRR* families expanded, seemingly independently, yet presumably driven by the distinct pathogen pressure faced by each species; this led to substantial variation in the size of different *NBS-LRR* families across species. This variation in copy number likely results from selection to maintain a diverse set of *NBS-LRR* families coupled with duplication and divergence within each *NBS-LRR* family—in both cases, this diversification yields distinct and advantageous resistance specificities. From our analysis of both presence/absence polymorphisms and the variation in copy number of *NBS-LRR* families, we can discern three evolutionary paths for *NBS-LRR* genes (fig. 5A). Focusing on *NBS-LRR* family evolution and setting aside the discussion of miRNA emergence: the first path is that an *NBS* gene stays as a single or low-copy gene, or it may be lost from the genome; a second path is the direct duplication of that *NBS-LRR* gene, forming potentially complex families; the third pathway is significant diversification of some *NBS-LRR*s, yielding new *NBS-LRR* families (fig. 5A). All three paths may be utilized in a given genome, for different families of *NBS-LRR* genes, and a relatively high heterogeneity of *NBS-LRR* genes may be indicative of a greater utilization of the third path. We observed variation in the level of heterogeneity of *NBS-LRR*s among different plant families, with the highest level in the Poaceae (fig. 1); the reason for this distinction remains unclear, perhaps resulting from a higher rate of diversification of the *NBS-LRR* families (i.e., the third path).

Despite this heterogeneity, some *NBS-LRR*s are highly retained across different species, including gene families that

may have evolved via the first two paths in our model. For example, the two most-conserved *NBS-LRR* families in our analysis were those encoding ADR1 and RPM1. ADR1 belongs to a special lineage of CNLs, RPW8-NBS-LRRs (RNLs), in which the N terminus contains an RPW8-like domain (Collier et al. 2011). ADR1 functions as a helper protein for other NBS-LRRs (Bonardi et al. 2011) and it can increase tolerance to abiotic stress (Chini et al. 2004). Our data show that the ADR1 family largely evolved via the first path, as a low-copy gene, except in the spruce genome. RPM1 functions to monitor the stability of RIN4 (Mackey et al. 2002), and its homologs were also present in both eudicots and monocots. The RPM1 family followed either the first or second paths in different plant families. The preservation of both the ADR1 and RPM1 families might be related to their functions, perhaps as key interacting "hubs" in the plant resistance gene signaling network.

Tandem duplication events, important for the second and third paths by which we envision *NBS-LRRs* evolve, are a major contributor to the expansion of a *NBS-LRR* families. These events are evident in the remarkably large clusters occasionally observed in some genomes, such as the 30+ *NBS-LRR* copies in the lettuce *Dm3* locus (Michelmore and Meyers 1998; Kuang et al. 2004). Direct tandem duplications may facilitate unequal crossing-over and family expansion (Luo et al. 2012), whereas heterogeneous clusters may be less likely to recombine. Some clusters of *NBS-LRR* genes are comprised of a special arrangement, as heterogeneous pairs in a divergent orientation (Meyers et al. 2003); these pairs have turned out to have functional importance because in some cases, they define pairs of proteins that function together to confer resistance (Williams et al. 2014; Saucet et al. 2015). We found that in the Poaceae and Brassicaceae, the number of heterogeneous *NBS-LRR* pairs is significantly higher than in other plant families.

The Evolutionary Tension and Balance of *NBS-LRR* Genes and the miRNAs That Target Them

miRNAs targeting *NBS-LRR* genes are frequently generated *de novo* from highly duplicated *NBS-LRR* genes. The targets of these miRNAs are, of course, their genic progenitors. Thus, the spontaneous generation of miRNAs might be driven by the expansion of *NBS-LRR* genes. Tandem duplication is a common mechanism by which *NBS-LRR* gene clusters expand (Michelmore and Meyers 1998); a localized rearrangement within such a cluster would yield a convergent or divergent pair. The classical model of miRNA emergence is that unknown genetic mechanisms could form an inverted repeat (a hairpin structure), the precursor to an miRNA (Allen et al. 2004). In the case of *NBS-LRRs*, we believe that inverted repeats are formed via unequal crossing over within clusters of these genes. The abundance of miRNAs targeting *NBS-LRRs* found across plant genomes supports several possible reasons that *NBS-LRRs* are exceptional sources of miRNAs: (1) they are frequently found as convergent or divergent pairs, so a relatively small rearrangement could form an inverted repeat; (2) there are many *NBS-LRRs* in plant genomes, creating many opportunities for rearrangements; (3) high conservation of the DNA sequence that encodes the P-loop means that even a rearrangement involving two very diverse *NBS-LRRs* could yield a short stem-loop structure. We observed a relationship between the emergence of new miRNAs and their regulation of the very genes from which they were derived; this represents a sort of evolutionary feedback loop (fig. 5A). Particularly notable was our observation that multiple miRNAs have emerged independently, in often quite distinct plant lineages, and target the same encoded protein motif (fig. 3B). This seems a clear case of convergent evolution, not previously described for plant miRNAs. However, duplication and subsequent divergence of a pre-existing miRNA precursor can also explain the fact that most miRNA target the P-loop, considering the ancient miR482/2118 family (Vries et al. 2015). At this point, it appears that the extent of the evolutionary relationship between miRNAs and *NBS-LRRs* is complex and dynamic (Vries et al. 2015), perhaps more than any other miRNA–target pair in plants, although aspects of this complexity can be seen in miRNAs and genic targets that encode MYB transcription factors, F-box proteins, and PPRs (Fei et al. 2013; Xia et al. 2015)

Though it is hypothesized that in the long term, miRNAs balance the benefits and costs of the duplication and diversification of *NBS-LRRs*, our results indicate that miRNAs may emerge in parallel to the expansion of *NBS-LRR* genes in a given plant lineage. There are genomic mechanisms by which the functional copies of *NBS-LRRs* can be minimized to reduce the "cost" of these genes (Tian et al. 2003). For example, a loss of function (i.e. pseudogenization) or expression would significantly reduce the genomic pool of active *NBS-LRRs*. Given that these miRNAs might not be critical or essential to plant survival (Boccara et al. 2014) and there is no strong selection acting on the sites for miRNAs, why then has the evolution of plant genomes independently and consistently resulted in miRNAs that regulate *NBS-LRRs*? miRNAs are a particularly powerful mechanism by which to modulate the transcript abundance of families of duplicated *NBS-LRRs*—one miRNA can function efficiently to regulate many *NBS-LRRs*, particularly when they target conserved motifs, functioning as 'master regulators' as in the case of the five Medicago miRNAs that target >100 *NBS-LRRs* (Zhai et al. 2011). It takes relatively few evolutionary events to yield a miRNA that regulates numerous targets, making it more "economical" to achieve a broad effect on the complement of active *NBS-LRRs* in a genome than the optimization of regulatory sequences across many genes (like promoters), or the large-scale accumulation of simple mutations (like pseudogenization).

Considering that miRNAs are a very efficient way to regulate duplicated *NBS-LRRs* as a group, from an evolutionary perspective, we might expect selection to drive the continual emergence of miRNAs that target *NBS-LRRs*. With this regulation taking place by varied miRNAs, there would be fewer constraints on the diversity of P-loop-encoding sequences. Indeed, this is what we observed: in eudicots, there was scant evidence for selection driven by the miRNA when we examined (1) the nucleotide diversity at the core of the P-loop, (2) the accessibility of the miR482/2118 target sites, and (3) the proportion of non-

synonymous substitutions. Instead, by examining how miRNAs varied at positions corresponding to their translated targets mRNAs, we found that the third position of the codon was highly variable in the miRNAs, indicating that these miRNAs have diversified to keep pace with their protein-coding targets. Intriguingly, one unexpected finding from our work was that the Brassicaceae and Poaceae are unusual families, as their genomes essentially lack the widespread targeting of *NBS-LRRs* by miR482/2118, and these families are distinguished by other characteristics of their *NBS-LRRs*. Thus, with the exception of these two unusual families, our data indicate that NBS-LRR protein function spurs miRNA diversification; however, it is less clear the extent to which miRNA–target pairing restricts NBS-LRR diversification—a topic for future investigation (fig. 5B).

Materials and Methods

Identification of Paralog/Ortholog Groups of Plant NBS-Encoding Genes

The genome sequences and the annotations of 75 plants were downloaded from Phytozome (http://phytozome.jgi.doe.gov/, v10, last accessed August 7, 2016) and other plant genome website (supplementary table S1, Supplementary Material online). The annotated proteins of each genome were searched against a model of the NBS domain (Pfam PF00931) to identify NBS-encoding candidates by hmmer3.0 (http://hmmer.janelia.org/, last accessed August 7, 2016) with an *E*-value cut-off of 10^{-4} (Yue et al. 2012). "NBS" most frequently means a portion of an NBS-LRR protein; we used "NBS" when presenting the data in the results and "NBS-LRR" to refer to the full protein when discussing their function and evolution. All the NBS-encoding candidates were searched against themselves using BLASTP with an *E*-value cut-off of 10^{-5} and then clustered into groups according to their pairwise normalized score using orthoMCL with default parameters (Li et al. 2003). The proteins of each group were re-annotated by NLR-parser (Steuernagel et al. 2015) and Interproscan (Quevillon et al. 2005) to obtain their domain information. Based on the presence/absence of an N-terminal (TIR, CC, RPW8) or C-terminal domain (LRR) in each NBS protein, the NBS candidates were classified into eight different types, namely the following: TIR-NBS-LRR (TNL), TIR-NBS (TN), CC-NBS-LRR (CNL), CC-NBS (CN), RPW8-NBS-LRR (RNL), RPW8-NBS (RN), NBS-LRR (NL), and NBS (N). We manually checked the groups of proteins which were not reported to contain NBS domain and we removed false positives, since some of proteins from the STAND family have considerable similarity with the NBS domain, as assessed by HMM searches (Luo et al. 2012). According to the protein groups, the gene IDs of all the NBS-encoding candidates were retrieved and set as paralog/ortholog gene families (*NBS* families) for further analysis. All the functional *R* genes in PRGdb (http://prgdb.crg.eu/, last accessed August 7, 2016) were mapped into the families using BLAST and used as the reference gene of the family to which they matched (Sanseverino et al. 2013).

Copy Number Variation, Genome Distribution, and Phylogeny of *NBS* Genes

The presence/absence and copy number variation of *NBS* genes in each family were calculated by customized Perl scripts. If two *NBS* genes are separated by no more than eight other non-*NBS* genes, they are considered to be located at the same *NBS* gene cluster (Richly et al. 2002). If a cluster contains *NBS* genes from multiple different families, it was defined as a heterogeneous *NBS* cluster. If a cluster only contains *NBS* genes from one family, the cluster was defined as a homogeneous *NBS* cluster. Two adjacent *NBS* genes (not separated by other genes) in a cluster were considered as an *NBS* gene pair. Their orientations in the genome were retrieved from GFF files of each genome. If a pair contained *NBS* genes from two different families, the pair was regarded as a heterogeneous NBS pair. *NBS* sequences were aligned using MUSCLE (Edgar 2004) and manually modified in GeneDoc (http://www.nrbsc.org/old/gfx/genedoc/, last accessed August 7, 2016). Neighbor-joining trees were constructed and bootstrap values were calculated using MEGA 6.0 (Tamura et al. 2013).

Prediction and Validation of miRNAs Targeting *NBS* Genes

All plant miRNA hairpins and mature sequences were downloaded from miRBase (Kozomara and Griffiths-Jones 2014) or prior publications. MicroRNAs targeting *NBS* genes were predicted by TargetFinder (Allen et al. 2005) and the target-matching scores were retrieved. Flanking sequence accessibility of miRNA target sites (upstream 17 nt and downstream 15 nt) was calculated by RNAup from the ViennaRNA Package 2.0 (Lorenz et al. 2011). MicroRNA–*NBS* interactions were selected when the matching score was ≤ 4 and the target accessibility was < -20 kcal/mol. miRNAs reportedly targeting *NBS* genes in *Arabidopsis*, Medicago, soybean and spruce were retrieved from previous studies (Zhai et al. 2011; Arikit et al. 2014; Boccara et al. 2014; Xia et al. 2015). Results in tomato and Brachypodium were validated visually in our genome browser using a phasing score >10 for target-derived 21-nt siRNAs (Nakano et al. 2006; Zhai et al. 2011).

Evolutionary Analysis of miRNAs and Their Target Sites

The presence/absence of miRNA families was designated in a phylogenetic tree to assess their loss and gain history (see main text). All the precursors of *NBS*-targeting miRNAs were searched against a database of the *NBS* sequences that we'd retrieved, using BLAST with an *E*-value cutoff of 10^{-5} to identify homologous genes. MicroRNA hairpins which matched *NBS* genes were also searched against NCBI nr database to confirm that the best-hit genes were *NBS-LRRs* and not other genes. For example, for miRNA aly-miR472, we first search the aly-miR472 hairpin sequence against our *NBS* database using BLASTN with the *E*-value cutoff set at 10^{-5}. The best hit of the aly-miR472 5′ and 3′ arms was *siri*|*NBS0045* and *alyr*|*NBS0176*, respectively. To make sure the best-matched coding sequence was *NBS-LRRs* but not other genes, we searched the aly-miR472 hairpin against NCBI using BLASTN with default parameters. The best

hit coding gene was NM_001160857.1, which is annotated as an *NBS-LRR*. Via this analysis, we ensured that the origin of alymiR472 was an *NBS-LRR*, and we repeated this for other miRNAs. The inverted repeats around *NBS-LRRs* in apple, cotton, Medicago, and tomato were extracted from the genome sequences. The secondary structure of the inverted repeat connecting two pairs of *NBS* genes was predicted using RNAfold (Lorenz et al. 2011).

Encoded P-loop motif sequences were extracted from proteins sequences derived from each *NBS* gene. The pairwise proportions of synonymous (P_s) and nonsynonymous (P_n) differences were calculated by SNAP (http://www.hiv.lanl.gov, last accessed August 7, 2016), for which the average was the synonymous and nonsynonymous nucleotide diversity defined in DnaSP (Librado and Rozas 2009). Sequences of miRNAs targeting *NBS* genes from each family were aligned individually and their diversified positions were detected manually. All the statistics analyses including descriptive analysis, *t*-test, and linear regression, were performed using the R package (https://www.r-project.org/, last accessed August 7, 2016).

Supplementary Material

Supplementary data S1–S3, tables S1–S11 and figures S1–S3 are available at *Molecular Biology and Evolution* online (http://www.mbe.oxfordjournals.org/).

Acknowledgments

We thank members of the Meyers lab for helpful discussions and input. This study was supported by US National Science Foundation, Division of Integrative Organismal Systems award #1257869, and a scholarship to Y. Zhang from the Chinese Scholarship Council.

References

Allen E, Xie Z, Gustafson AM, Carrington JC. 2005. microRNA-directed phasing during *trans*-acting siRNA biogenesis in plants. *Cell* 121:207–221.

Allen E, Xie Z, Gustafson AM, Sung G-H, Spatafora JW, Carrington JC. 2004. Evolution of microRNA genes by inverted duplication of target gene sequences in *Arabidopsis thaliana*. *Nat Genet*. 36:1282–1290.

Arikit S, Xia R, Kakrana A, Huang K, Zhai J, Yan Z, Valdés-López O, Prince S, Musket TA, Nguyen HT, et al. 2014. An atlas of soybean small RNAs identifies phased siRNAs from hundreds of coding genes. *Plant Cell* 26:4584–4601.

Axtell MJ. 2013. Classification and comparison of small RNAs from plants. *Annu Rev Plant Biol*. 64:137–159.

Boccara M, Sarazin A, Thiebeauld O, Jay F, Voinnet O, Navarro L, Colot V. 2014. The Arabidopsis miR472-RDR6 silencing pathway modulates PAMP- and effector-triggered immunity through the post-transcriptional control of disease resistance genes. *PLoS Pathog*. 10:e1003883.

Bonardi V, Tang S, Stallmann A, Roberts M, Cherkis K, Dangl JL. 2011. Expanded functions for a family of plant intracellular immune receptors beyond specific recognition of pathogen effectors. *Proc Natl Acad Sci U S A*. 108:16463–16468.

Chen H-M, Chen L-T, Patel K, Li Y-H, Baulcombe DC, Wu S-H. 2010. 22-Nucleotide RNAs trigger secondary siRNA biogenesis in plants. *Proc Natl Acad Sci U S A*. 107:15269–15274.

Chini A, Grant JJ, Seki M, Shinozaki K, Loake GJ. 2004. Drought tolerance established by enhanced expression of the CC-NBS-LRR gene, ADR1, requires salicylic acid, EDS1 and ABI1. *Plant J*. 38:810–822.

Collier SM, Hamel LP, Moffett P. 2011. Cell death mediated by the N-terminal domains of a unique and highly conserved class of NB-LRR protein. *Mol Plant Microbe Interact*. 24:918–931.

Cuperus JT, Carbonell A, Fahlgren N, Garcia-Ruiz H, Burke RT, Takeda A, Sullivan CM, Gilbert SD, Montgomery TA, Carrington JC. 2010. Unique functionality of 22-nt miRNAs in triggering RDR6-dependent siRNA biogenesis from target transcripts in Arabidopsis. *Nat Struct Mol Biol*. 17:997–1003.

Edgar RC. 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acid Res*. 32:1792–1797.

Eitas TK, Dangl JL. 2010. NB-LRR proteins: pairs, pieces, perception, partners, and pathways. *Curr Opin Plant Biol*. 13:472–477.

Fei Q, Li P, Teng C, Meyers BC. 2015. Secondary siRNAs from *Medicago NB-LRRs* modulated via miRNA-target interactions and their abundances. *Plant J*. 83:451–465.

Fei Q, Xia R, Meyers BC. 2013. Phased, secondary, small interfering RNAs in posttranscriptional regulatory networks. *Plant Cell* 25:2400–2415.

Fei Q, Zhang Y, Xia R, Meyers BC. 2016. Small RNAs add zing to the zig-zag-zig model of plant defenses. *Mol Plant Microbe Interact*. 29:165–169.

Felippes FFD, Schneeberger K, Dezulian T, Huson DH, Weigel D. 2008. Evolution of *Arabidopsis thaliana* microRNAs from random sequences. *RNA* 14:2455–2459.

González VM, Müller S, Baulcombe D, Puigdomènech P. 2015. Evolution of NBS-LRR gene copies among dicot plants and its regulation by members of the miR482/2118 superfamily of miRNAs. *Mol Plant* 8:329–331.

He XF, Fang YY, Feng L, Guo HS. 2008. Characterization of conserved and novel microRNAs and their targets, including a TuMV-induced TIR-NBS-LRR class R gene-derived novel miRNA in Brassica. *FEBS Lett*. 582:2445–2452.

Johnson C, Kasprzewska A, Tennessen K, Fernandes J, Nan GL, Walbot V, Sundaresan V, Vance V, Bowman LH. 2009. Clusters and superclusters of phased small RNAs in the developing inflorescence of rice. *Genome Res*. 19:1429–1440.

Jones JDG, Dangl JL. 2006. The plant immune system. *Nature* 444:323–329.

Kertesz M, Iovino N, Unnerstall U, Gaul U, Segal E. 2007. The role of site accessibility in microRNA target recognition. *Nat Genet*. 39:1278–1284.

Kozomara A, Griffiths-Jones S. 2014. miRBase: annotating high confidence microRNAs using deep sequencing data. *Nucleic Acids Res*. 42:D68–D73.

Krom N, Ramakrishna W. 2008. Comparative analysis of divergent and convergent gene pairs and their expression patterns in rice, Arabidopsis, and populus. *Plant Physiol*. 147:1763–1773.

Kuang H, Woo SS, Meyers BC, Nevo E, Michelmore RW. 2004. Multiple genetic processes result in heterogeneous rates of evolution within the major cluster disease resistance genes in lettuce. *Plant Cell* 16:2870–2894.

Li F, Pignatta D, Bendix C, Brunkard JO, Cohn MM, Tung J, Sun H, Kumar P, Baker B. 2012. MicroRNA regulation of plant innate immune receptors. *Proc Natl Acad Sci U S A*. 109:1790–1795.

Li L, Roos DS, Stoeckert CJ Jr. 2003. OrthoMCL: identification of ortholog groups for eukaryotic genomes. *Genome Res*. 2003:2178–2189.

Librado P, Rozas J. 2009. DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. *Bioinformatics* 25:1451–1452.

Lin X, Zhang Y, Kuang H, Chen J. 2013. Frequent loss of lineages and deficient duplications accounted for low copy number of disease resistance genes in Cucurbitaceae. *BMC Genomics* 14:335.

Liu J, Cheng X, Liu D, Xu W, Wise R, Shen Q-H. 2014. The miR9863 family regulates distinct Mla alleles in barley to attenuate NLR receptor-triggered disease resistance and cell-death signaling. *PLoS Genet*. 10:e1004755.

Lorenz R, Bernhart SH, Höner Zu Siederdissen C, Tafer H, Flamm C, Stadler PF, Hofacker IL. 2011. ViennaRNA Package 2.0. *Algorithm Mol Biol*. 6:26.

Lu C, Kulkarni K, Souret FF, MuthuValliappan R, Tej SS, Poethig RS, Henderson IR, Jacobsen SE, Wang W, Green PJ, et al. 2006. MicroRNAs and other small RNAs enriched in the Arabidopsis RNA-dependent RNA polymerase-2 mutant. *Genome Res*. 16:1276–1288.

Luo S, Zhang Y, Hu Q, Chen J, Li K, Lu C, Liu H, Wang W, Kuang H. 2012. Dynamic nucleotide-binding site and leucine-rich repeat-encoding genes in the grass family. *Plant Physiol*. 159:197–210.

Mackey D, Holt BF, Wiig A, Dangl JL. 2002. RIN4 interacts with *Pseudomonas syringae* type III effector molecules and is required for RPM1-mediated resistance in Arabidopsis. *Cell* 108:743–754.

Maekawa T, Kufer TA, Schulze-Lefert P. 2011. NLR functions in plant and animal immune systems: so far and yet so close. *Nat Immunol*. 12:817–826.

Meyers BC, Dickerman AW, Michelmore RW, Sobral BW. 1999. Plant disease resistance genes encode members of an ancient and diverse protein family within the nucleotide-binding superfamily. *Plant J*. 20:317–332.

Meyers BC, Kozik A, Griego A, Kuang H, Michelmore RW. 2003. Genome-wide analysis of NBS-LRR-encoding genes in Arabidopsisnucleotide-binding superfamily. *Plant Cell*. 15:809–834.

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

Monaghan J, Zipfel C. 2012. Plant pattern recognition receptor complexes at the plasma membrane. *Curr Opin Plant Biol*. 15:349–357.

Mondragon-Palomino M, Meyers BC, Michelmore RW, Gaut BS. 2002. Patterns of positive selection in the complete NBS-LRR gene family of *Arabidopsis thaliana*. *Genome Res*. 12:1305–1315.

Mukhtar MS, Carvunis AR, Dreze M, Epple P, Steinbrenner J, Moore J, Tasan M, Galli M, Hao T, Nishimura MT, et al. 2011. Independently evolved virulence effectors converge onto hubs in a plant immune system network. *Science* 333:596–601.

Nakano M, Nobuta K, Vemaraju K, Tej SS, Skogen JW, Meyers BC. 2006. Plant MPSS databases: signature-based transcriptional resources for analyses of mRNA and small RNA. *Nucleic Acids Res*. 34:D731–D735.

Ouyang S, Park G, Atamian HS, Han CS, Stajich JE, Kaloshian I, Borkovich KA. 2014. MicroRNAs suppress NB domain genes in tomato that confer resistance to *Fusarium oxysporum*. *PLoS Pathog*. 10:e1004464.

Quevillon E, Silventoinen V, Pillai S, Harte N, Mulder N, Apweiler R, Lopez R. 2005. InterProScan: protein domains identifier. *Nucleic Acids Res*. 33:116–120.

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.

Sanseverino W, Hermoso A, D'Alessandro R, Vlasova A, Andolfo G, Frusciante L, Lowy E, Roma G, Ercolano MR. 2013. PRGdb 2.0: towards a community-based database model for the analysis of R-genes in plants. *Nucleic Acids Res*. 41:D1167–D1171.

Saucet SB, Ma Y, Sarris PF, Furzer OJ, Sohn KH, Jones JDG. 2015. Two linked pairs of Arabidopsis TNL resistance genes independently confer recognition of bacterial effector AvrRps4. *Nat Commun*. 6:6338.

Shao Z-Q, Zhang Y-M, Hang Y-Y, Xue J-Y, Zhou G-C, Wu P, Wu X-Y, Wu X-Z, Wang Q, Wang B, Chen J. 2014. Long-term evolution of nucleotide-binding site-leucine-rich repeat (NBS-LRR) genes: understandings gained from and beyond the legume family. *Plant Physiol*. 166:217–234.

Shivaprasad PV, Chen HM, Patel K, Bond DM, Santos BA, Baulcombe DC. 2012. A microRNA superfamily regulates nucleotide binding site-leucine-rich repeats and other mRNAs. *Plant Cell* 24:859–874.

Steuernagel B, Jupe F, Witek K, Jones JDG, Wulff BH. 2015. NLR-parser: rapid annotation of plant NLR complements. *Bioinformatics* 31:1665–1667.

Stokes TL, Kunkel BN, Richards EJ. 2002. Epigenetic variation in Arabidopsis disease resistance. *Genes Dev*. 16:171–182.

Takken FL, Tameling WI. 2009. To nibble at plant resistance proteins. *Science* 324:744–746.

Tamura K, Stecher G, Peterson D, Filipski A, Kumar S. 2013. MEGA6: molecular evolutionary genetics analysis version 6.0. *Mol Biol Evol*. 30:2725–2729.

Tarr DE, Alexander HM. 2009. TIR-NBS-LRR genes are rare in monocots: evidence from diverse monocot orders. *BMC Res Notes* 2:197.

Tian D, Traw MB, Chen JQ, Kreitman M, Bergelson J. 2003. Fitness costs of R-gene-mediated resistance in *Arabidopsis thaliana*. *Nature* 423:74–77.

Vries S, Kloesges T, Rose LE. 2015. Evolutionarily dynamic, but robust, targeting of resistance genes by the miR482/2118 gene family in the Solanaceae. *Genome Biol Evol*. 7:3307–3321.

Wang J, Tan S, Zhang L, Li P, Tian D. 2011. Co-variation among major classes of LRR-encoding genes in two pairs of plant species. *J Mol Evol*. 72:498–509.

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

Williams SJ, Sohn KH, Wan L, Bernoux M, Sarris PF, Segonzac C, Ve T, Ma Y, Saucet SB, Ericsson DJ, et al. 2014. Structural basis for assembly and function of a heterodimeric plant immune receptor. *Science* 344:299–303.

Xia R, Meyers BC, Liu Z, Beers EP, Ye S, Liu Z. 2013. MicroRNA superfamilies descended from miR390 and their roles in secondary small interfering RNA biogenesis in eudicots. *Plant Cell* 25:1555–1572.

Xia R, Xu J, Arikit S, Meyers BC. 2015. Extensive families of miRNAs and *PHAS* loci in Norway spruce demonstrate the origins of complex phasiRNA networks in seed plants. *Mol Biol Evol*. 32:2905–2918.

Xia R, Ye S, Liu Z, Meyers BC, Liu Z. 2015. Novel and recently evolved microRNA clusters regulate expansive F-BOX gene networks through phased small interfering RNAs in wild diploid strawberry. *Plant Physiol*. 169:594–610.

Yue JX, Meyers BC, Chen JQ, Tian D, Yang S. 2012. Tracing the origin and evolutionary history of plant nucleotide-binding site-leucine-rich repeat (NBS-LRR) genes. *New Phytol*. 193:1049–1063.

Zhai J, Jeong DH, De Paoli E, Park S, Rosen BD, Li Y, Gonzalez AJ, Yan Z, Kitto SL, Grusak MA, et al. 2011. MicroRNAs as master regulators of the plant NB-LRR defense gene family via the production of phased, *trans*-acting siRNAs. *Genes Dev*. 25:2540–2553.

Zhai J, Zhang H, Arikit S, Huang K, Nan G-L, Walbot V, Meyers BC. 2015. Spatiotemporally dynamic, cell-type–dependent premeiotic and meiotic phasiRNAs in maize anthers. *Proc Natl Acad Sci*. 112:3146–3151.

Zhang R, Murat F, Pont C, Langin T, Salse J. 2014. Paleo-evolutionary plasticity of plant disease resistance genes. *BMC Genomics* 15:187.

Zhao M, Meyers BC, Cai C, Xu W, Ma J. 2015. Evolutionary patterns and coevolutionary consequences of MIRNA genes and microRNA targets triggered by multiple mechanisms of genomic duplications in soybean. *Plant Cell* 27:546–562.

Zhu H, Guo H. 2012. The role of virus-derived small interfering RNAs in RNA silencing in plants. *Sci China Life Sci*. 55:119–125.