Large-Scale Analyses of Angiosperm Nucleotide-Binding Site-Leucine-Rich Repeat Genes Reveal Three Anciently Diverged Classes with Distinct Evolutionary Patterns

Zhu-Qing Shao2, Jia-Yu Xue2, Ping Wu, Yan-Mei Zhang, Yue Wu, Yue-Yu Hang, Bin Wang*, and Jian-Qun Chen*

State Key Laboratory of Pharmaceutical Biotechnology, School of Life Sciences, Nanjing University, Nanjing, 210023, China (Z.-Q.S., P.W., Y.-M.Z., Y.W., B.W., J.-Q.C.); and Institute of Botany, Jiangsu Province and Chinese Academy of Sciences, Nanjing, 210014, China (J.-Y.X., Y.-M.Z., Y.-Y.H.)

Nucleotide-binding site-leucine-rich repeat (NBS-LRR) genes make up the largest plant disease resistance gene family (R genes), with hundreds of copies occurring in individual angiosperm genomes. However, the expansion history of NBS-LRR genes during angiosperm evolution is largely unknown. By identifying more than 6,000 NBS-LRR genes in 22 representative angiosperms and reconstructing their phylogenies, we present a potential framework of NBS-LRR gene evolution in the angiosperm. Three anciently diverged NBS-LRR classes (TNLs, CNLs, and RNLs) were distinguished with unique exon-intron structures and DNA motif sequences. A total of seven ancient TNL, 14 CNL, and two RNL lineages were discovered in the ancestral angiosperm, from which all current NBS-LRR gene repertoires were evolved. A pattern of gradual expansion during the first 100 million years of evolution of the angiosperm clade was observed for CNLs. TNL numbers remained stable during this period but were eventually deleted in three divergent angiosperm lineages. We inferred that an intense expansion of both TNL and CNL genes started from the Cretaceous-Paleogene boundary. Because dramatic environmental changes and an explosion in fungal diversity occurred during this period, the observed expansions of R genes probably reflect convergent adaptive responses of various angiosperm families. An ancient whole-genome duplication event that occurred in an angiosperm ancestor resulted in two RNL lineages, which were conservatively evolved and acted as scaffold proteins for defense signal transduction. Overall, the reconstructed framework of angiosperm NBS-LRR gene evolution in this study may serve as a fundamental reference for better understanding angiosperm NBS-LRR genes.

Plant disease resistance (R) genes are a set of genes that confer resistance to various pathogens. Among all known types of R genes, nucleotide-binding site-leucine-rich repeat (NBS-LRR, NBS for short) genes represent the largest class, encompassing more than 80% of the characterized R genes (Meyers et al., 1999; Meyers et al., 2005; McHale et al., 2006; Friedman and Baker, 2007). Since the first comprehensive study on NBS genes in Arabidopsis (Arabidopsis thaliana; Meyers et al., 2003), NBS genes have been surveyed across various plant genomes, most of which belong to the rosid lineage of eudicots and the Poaceae of monocots (Monosi et al., 2004; Zhou et al., 2004; Yang et al., 2006; Ameline-Torregrosa et al., 2008; Yang et al., 2008; Mun et al., 2009; Porter et al., 2009; Li et al., 2010; Zhang et al., 2011; Lozano et al., 2012; Luo et al., 2012; Andolfi et al., 2013; Shao et al., 2014).

NBS genes are historically divided into two classes, namely, TIR-NBS-LRR (TNL) and the non-TIR-NBS-LRR (nTNL), which are differentiated by the presence of a Toll/IL-1 Receptor-like (TIR) domain in the protein amino terminus (Meyers et al., 1999; Bai et al., 2002; Zhou et al., 2004). Because most nTNL genes encode a coiled-coil (CC) domain at the N terminus, the nTNL genes often are called CC-NBS-LRR (CNL) genes (Meyers et al., 2003; Ameline-Torregrosa et al., 2008). However, recent studies have revealed that apart from CNL genes, a small group of nTNL genes that possess a special N-terminal domain, RPW8 (resistance to powdery mildew8) domain, likely represent a distinct class of NBS genes (RPW8-NBS-LRR [RNL]; Xiao et al., 2001; Cannon et al., 2004; Bonardi et al., 2011; Collier et al., 2011; Shao et al., 2014; Zhang et al., 2016). Investigations of NBS genes in the Fabaceae (rosid I lineage) and...
Brassicaceae (rosid II lineage) have shown that RNLs and CNLs diverged prior to the divergence of these two families (Shao et al., 2014; Zhang et al., 2016). Interestingly, although both CNL and RNL genes are present in monocots and dicots, TNL genes only are present in dicots (Meyers et al., 2003; Tarr and Alexander, 2009; Andolfo et al., 2013; Shao et al., 2014). Considering the abundance of TNLs in various dicot species, their absence in monocots is perplexing.

For several years, our knowledge of NBS gene evolution has been based on investigations of individual angiosperm genomes (Meyers et al., 2003; Ameline-Torregrosa et al., 2008; Kohler et al., 2008). The number of NBS genes varied dramatically among different species, even among different accessions of the same species (Zhang et al., 2010). However, the relative contributions of recent and ancient expansions to the abundance of NBS genes in modern plant genomes remain elusive. Recent studies report the evolutionary analyses of NBS genes across multiple species in a plant family (Cannon et al., 2002; Guo et al., 2011; Luo et al., 2012; Lin et al., 2013; Shao et al., 2014; Zhang et al., 2016). While studies in Brassicaceae and Poaceae both revealed abundant and recent losses of NBS genes in various species after the divergence of these two families, a study involving Fabaceae genome data has suggested the opposite scenario in which all surveyed species recently underwent a multitude of NBS expansions (Luo et al., 2012; Shao et al., 2014; Zhang et al., 2016). Although distinct evolutionary patterns of NBS genes have been identified at the family level, hundreds of NBS gene lineages arose in the ancestors of Fabaceae, Brassicaceae, and Poaceae, indicating that NBS genes may have probably undergone ancient expansions prior to familial divergence. Therefore, a phylogenetic analysis at a larger scale is necessary to better understand the expansion history of NBS genes in angiosperms.

Several functional NBS genes have been identified in angiosperm species (Liu et al., 2007); however, their time of emergence and relationship with other NBS genes have not been established. For example, several CNL genes have been identified in both soybean (Glycine max; Rpg1b and Rpg1r) and Arabidopsis (RPM1 and RPS2) against the bacterial pathogen Pseudomonas syringae by guarding the host protein RIN4 (Bent et al., 1994; Mackey et al., 2002; Mackey et al., 2003; Ashfield et al., 2014). While the Rpg1b and Rpg1r genes were deduced to emerge after the divergence of Fabaceae species, the evolutionary time point of the origin of RPM1 and RPS2 genes in Arabidopsis is largely unknown (Ashfield et al., 2014; Shao et al., 2014). Because all these R genes have likely adopted a similar strategy to defend against the same pathogen, investigating their evolution may elucidate the mechanism underlying the establishment and maintenance of resistance in various plant species.

To obtain a more comprehensive understanding of NBS gene evolution, it is essential to analyze more species/lineages to determine how the current NBS genes have evolved and diverged from ancestral NBS lineages in the ancestral angiosperm. This will provide when and how resistance loci emerged or disappeared in certain lineages. In this study, a total of 22 plant species (Fig. 1) from key systematic clades in eudicots, monocots, and the early diverging basal angiosperm were chosen to depict the ancient evolutionary history of NBS genes in angiosperms. More than 6,000 NBS genes were identified; family level phylogenetic trees were constructed first and then expanded to higher systematic levels to eventually analyze all of the angiosperms. Ancient taxonomies of NBS genes were traced to the origin of angiosperms, and a total of 23 ancestral NBS genes were recovered in the last common ancestor of angiosperms. This angiosperm NBS frame provides a fundamental reference for understanding the evolutionary history of NBS genes as well as insight into the finer details of particular NBS gene classes.

RESULTS

Identification and Classification of NBS Genes from 22 Angiosperm Species

A total of 6,135 NBS genes were identified in 22 angiosperm species using previously described procedures (Shao et al., 2014; Zhang et al., 2016). Five of the 22 genomes were surveyed in this study, Amborella trichopoda, Musa acuminata (banana), Phyllostachys heterocycla (bamboo), Capsicum annum (hot pepper), and Sesamum indicum (sesame), whereas the others have been previously reported (Meyers et al., 2003; Luo et al., 2012; Malacarne et al., 2012; Andolfo et al., 2013; Shao et al., 2014; Zhang et al., 2016). A. trichopoda, a basal angiosperm, had a total of 105 NBS genes, including 15 TNLs, one RNL, and 89 CNLs (Fig. 1; Supplemental Table S1). The NBS genes varied among the 22 surveyed species (Fig. 1). Of all species surveyed, Medicago truncatula possessed the highest number of NBS genes (571), more than 6-fold of that observed in Thellungiella salsugina (88). RNL genes were detected in all of the surveyed species, although at lower numbers (1–18). A large proportion of NBS genes were either TNLs or CNLs, although TNLs were not detected in eight of 22 species, including seven monocots and S. indicum. Because the basal angiosperm A. trichopoda harbored TNL genes, these genes were most likely also present in the ancestor of angiosperms but independently lost in certain lineages. In species with both TNL and CNL genes, CNLs often outnumbered TNLs, except for the four Brassicaceae genomes analyzed (Zhang et al., 2016). The asterid genomes showed a lower ratio of TNL to CNL genes, with the C. annum exhibiting the highest level of disparity (TNL:CNL = 1:18).

Phylogenetic Analysis of NBS Genes in Major Angiosperm Clades

The 22 angiosperm species surveyed in this study occupy important phylogenetic positions (Fig. 1). To
trace the evolutionary history of NBS genes, we examined multiple angiosperm clades, including rosids (Ro), asterids (As), eudicots (Ed), monocots (Mo), mesangiospermae (Ma), and all angiosperms (An; Fig. 1). It is widely recognized that the NBS domain effectively distinguishes NBS genes into TNL and nTNL classes (Bai et al., 2002; Meyers et al., 2003). Although a sister relationship between CNLs and RNLs has been established in rosids, it has not been examined at angiosperm level. Therefore, phylogenetic analyses of TNLs and nTNLs (including both CNLs and RNLs) were separately performed in this study. We first conducted phylogenetic analysis of three major angiosperm lineages, namely, monocots, asterids, and rosids, which cover seven, four, and 10 genomes, respectively.

The *M. acuminata* genome, together with six Poaceae genomes, including *Oryza sativa*, *Brachypodium distachyon*, *P. heterocycla*, *Zea mays*, *Sorghum bicolor*, and *Setaria italica*, were used to elucidate the evolution of NBS genes in monocots. Because these species did not harbor TNL genes, phylogenetic analysis of only nTNL genes was performed. The reconstructed monocot nTNL phylogeny (Supplemental Fig. S1) revealed that nTNL genes from the six Poaceae species formed 456 monophyletic lineages, indicating the presence of at least 456 ancestral nTNL genes in the Poaceae common ancestor (~62 million years ago [MYA]; Peng et al., 2013, Zhang et al., 2012). These Poaceae nTNL genes formed 57 monophyletic lineages when grouped with *M. acuminata* nTNL genes (one RNL lineage designated as Mo-R1, and 56 CNL lineages identified as Mo-C1 to Mo-C56), suggesting that nTNL genes from the seven monocot genomes were actually derived from 57 ancestral Mo-NBS genes ~110 MYA (Fig. 2; Supplemental Fig. S1; D’Hont et al., 2012). From their common ancestor, the Poaceae and *M. acuminata* inherited 51 and 32 ancestral Mo-NBS genes, respectively. Among these, 26 were shared by both Poaceae and *M. acuminata*. Several ancestral genes in the Poaceae underwent extensive duplications (from 51 to 456) during the time interval of 48 million years. For example, the NBS gene Mo-C56 duplicated into 197 genes during this time period (Supplemental Fig. S1). As multiple species radiated from the Poaceae common ancestor, in some species NBS gene numbers contracted drastically (*Z. mays*, 139 from 456) and in others only slightly (*S. italica* and *P. heterocycla*, 424 and 344 from 456, respectively), whereas expansion was only found in the *O. sativa* genome (from 456 to 498).

Within the asterids four species were analyzed, including three Solanaceae species, namely, *Solanum lycopersicum*, *Solanum tuberosum*, and *C. annuum*, as well as *S. indicum* from Pedaliaceae. NBS genes from the three Solanaceae genomes formed 19 TNL monophyletic lineages (Supplemental Fig. S2) and 147 nTNL monophyletic lineages (Supplemental Fig. S3), suggesting that at least 166 ancestral NBS genes were present in their last common ancestor. The Solanaceae nTNL genes formed 55 monophyletic lineages with nTNL genes from *S. indicum* (two RNL lineages identified as Mo-C1 to Mo-C56), suggesting that nTNL genes from the seven monocot genomes were actually derived from 57 ancestral Mo-NBS genes ~110 MYA (Fig. 2; Supplemental Fig. S1; D’Hont et al., 2012). From their common ancestor, the Poaceae and *M. acuminata* inherited 51 and 32 ancestral Mo-NBS genes, respectively. Among these, 26 were shared by both Poaceae and *M. acuminata*. Several ancestral genes in the Poaceae underwent extensive duplications (from 51 to 456) during the time interval of 48 million years. For example, the NBS gene Mo-C56 duplicated into 197 genes during this time period (Supplemental Fig. S1). As multiple species radiated from the Poaceae common ancestor, in some species NBS gene numbers contracted drastically (*Z. mays*, 139 from 456) and in others only slightly (*S. italica* and *P. heterocycla*, 424 and 344 from 456, respectively), whereas expansion was only found in the *O. sativa* genome (from 456 to 498).

Within the asterids four species were analyzed, including three Solanaceae species, namely, *Solanum lycopersicum*, *Solanum tuberosum*, and *C. annuum*, as well as *S. indicum* from Pedaliaceae. NBS genes from the three Solanaceae genomes formed 19 TNL monophyletic lineages (Supplemental Fig. S2) and 147 nTNL monophyletic lineages (Supplemental Fig. S3), suggesting that at least 166 ancestral NBS genes were present in their last common ancestor. The Solanaceae nTNL genes formed 55 monophyletic lineages with nTNL genes from *S. indicum* (two RNL lineages identified as Mo-C1 to Mo-C56), suggesting that nTNL genes from the seven monocot genomes were actually derived from 57 ancestral Mo-NBS genes ~110 MYA (Fig. 2; Supplemental Fig. S1; D’Hont et al., 2012). From their common ancestor, the Poaceae and *M. acuminata* inherited 51 and 32 ancestral Mo-NBS genes, respectively. Among these, 26 were shared by both Poaceae and *M. acuminata*. Several ancestral genes in the Poaceae underwent extensive duplications (from 51 to 456) during the time interval of 48 million years. For example, the NBS gene Mo-C56 duplicated into 197 genes during this time period (Supplemental Fig. S1). As multiple species radiated from the Poaceae common ancestor, in some species NBS gene numbers contracted drastically (*Z. mays*, 139 from 456) and in others only slightly (*S. italica* and *P. heterocycla*, 424 and 344 from 456, respectively), whereas expansion was only found in the *O. sativa* genome (from 456 to 498).
designated as As-R1 and As-R2, and 53 CNL lineages identified as As-C1 to As-C53), indicating a minimum of 55 ancestral genes in the asterid common ancestor (Fig. 2; Supplemental Fig. S2). Solanales inherited more As-nTNL lineages than *S. indicum*, which represents Lamiales (47 versus 30). No TNL genes were detected in the *S. indicum* genome; therefore, the accurate number of ancestral TNL genes at this asterid node was not estimated.

Rosids are a major lineage of eudicots that consists of three sublineages, rosid I (fabids), rosid II (malvids), and the basal rosids. Previous work analyzed NBS genes in Fabaceae and Brassicaceae (Shao et al., 2014; Zhang et al., 2016), which belong to the rosid I sublineage and the rosid II sublineage, respectively. A total of 64 nTNL and 55 TNL ancestral lineages were discovered in the last common ancestor of the four
Fabaceae species, and 80 nTNL and 148 TNL ancestral lineages were identified in the last common ancestor of the five surveyed Brassicaceae species. In this study, one representative sequence from each of the 119 Fabaceae ancestral NBS lineages and the 228 Brassicaceae ancestral NBS lineages were used to represent the diversity of ancient NBS genes in these two families, and then combined with NBS genes identified in Vitis vinifera (basal rosid) to construct NBS phylogenies for the whole rosid lineage. A total of 61 ancestral Ro-NBS genes were identified (four RNLs, 49 CNLs, and eight TNLs) in the common ancestor of the rosids (~108 MYA; Fig. 2; Supplemental Figs. S4 and S5), among which the Fabaceae of rosid I lineage inherited 35 Ro-NBS genes (two RNLs, 27 CNLs, and six TNLs), Brassicaceae of rosid II lineage inherited 19 (two RNLs, 12 CNLs, and five TNLs), and basal rosids inherited 33 (two RNLs, 29 CNLs, and two TNLs). Less than half of the Ro-NBS genes (26/61) were detected in two of the lineages, and 35/61 were observed in one descending lineage (Fig. 2), thereby reflecting a pattern of frequent NBS gene loss after their split from the common rosid ancestor. The 35 and 19 Ro-NBS genes inherited by Fabaceae and Brassicaceae lineages expanded to 119 (55 TNLs and 64 nTNLs) and 228 (148 TNLs and 80 nTNLs) NBS genes in the Fabaceae ancestor (~54 MYA) and the Brassicaceae ancestor (~42 MYA), respectively (Shao et al., 2014; Zhang et al., 2016). The NBS genes continued to expand in Fabaceae (Shao et al., 2014), whereas these genes contracted as the Brassicaceae diverged into multiple species (Zhang et al., 2016). Although the Brassicaceae inherited fewer ancestral TNLs than CNLs (five versus 12) from the rosid ancestor, this lineage vigorously expanded its TNLs during the time interval of ~108 MYA to ~42 MYA. A 30-fold expansion was observed in the Brassicaceae TNLs (from five to 148), far beyond that of legume TNLs (~9-fold, from six to 55; Shao et al., 2014; Zhang et al., 2016).

Reconstruction of NBS Gene Phylogenies at the Whole-Angiosperm Scale

Representative NBS sequences from ancestral lineages of asterids, rosids, and monocots were used to construct ancestral NBS gene lineages at the divergence nodes of eudicots and mesangiospermae. A total of eight Ed-TNL and 49 Ed-nTNL genes (47 Ed-CNLS and two Ed-RNLs) were recovered at the eudicot node (Supplemental Figs. S6 and S7), whereas 37 Ma-nTNL genes (35 Ma-CNLS and two Ma-RNLs) were detected at the mesangiospermae node (Supplemental Fig. S8). Independent loss and differential inheritance of ancestral NBS genes were also observed at these two nodes. Representative genes from each ancestral Ma-nTNL lineage were compiled with A. trichopoda nTNL genes to reconstruct angiosperm nTNL phylogeny (Supplemental Fig. S9). Because a mesangiospermae TNL phylogeny could not be constructed due to the absence of TNL genes in monocots, a TNL phylogeny was generated at the angiosperm level using eudicot representative TNLs and A. trichopoda TNLs (Supplemental Fig. S10). Figure 3 shows that all TNL genes clustered into seven ancestral angiosperm lineages (designated as An-T1 to An-T7), and all nTNL genes clustered into 16 ancestral lineages (two RNL lineages designated as An-R1 and An-R2, and 14 CNL lineages identified as An-C1 to An-C14), indicating at least 23 NBS genes were present in the last angiosperm ancestor (~225 MYA). The basal angiosperm A. trichopoda inherited only eight of the 23 An-NBS genes, whereas the mesangiospermae inherited 19 (Fig. 3). The angiosperm nTNL tree was composed of two monophyletic groups, namely, RNL and CNL (Supplemental Fig. S9). Therefore, despite the extensive discrepancy in gene numbers between the two groups, this study revealed that the RNL is another ancient class of NBS genes that has a sister relationship with the CNL class in angiosperms.

The angiosperm NBS gene phylogenies have revealed the evolutionary state of NBS genes at critical angiosperm nodes, which indicated that differential inheritance of ancestral NBS genes and lineage-specific expansions are major themes of NBS gene evolution.

Characterization of Class-Specific Signatures among Three Ancient NBS Gene Classes

Because the results of phylogenetic analysis suggest that RNLs represent an ancient NBS gene class independent of the CNL and TNL classes, characterization of distinctive features of each class was performed.

Distinctive Signatures in Conserved NBS Motifs Differentiate the Three Classes

The NBS domain is composed of a number of characteristic motifs that are conserved and shared by all NBS genes, whereas other motifs are more variable (DeYoung and Innes, 2006). All three classes of NBS genes share five functionally important motifs: P-loop, Kinase-2, RNBS-B, GLPL, and RNBS-D. The amino acid sequence conservation of these five motifs was estimated using sequence logos using MEME software (Fig. 4; Bailey and Elkan, 1995). All motifs were highly conserved at most sites among all classes, except for the RNBS-D motif, which was poorly conserved in the TNL and CNL classes. The class-specific features within these motifs also were identified. The Kinase-2 motif had a conserved “DDVW” sequence in the RNL and CNL genes, but frequently appeared as “DDVD” in the TNL genes. A “TSR” sequence in the RNBS-B motif of RNLs was conserved, but frequently appeared as “TTR” and “TTRD” in the CNL and TNL genes, respectively. CNL did not have specific highly conserved motifs, but could be distinguished by a RNL like Kinase-2 motif and a TNL like RNBS-B motif.

Intron-Exon Structures Also Distinguish NBS Classes

The absence/presence, position, and phase of introns often are regarded as important parameters in phylogeny (Qiu et al., 1998; Sánchez et al., 2003; Qian et al.,
In a previous study, the intron positions and phases were determined to be distinctive among the classes of NBS genes in bryophytes (Xue et al., 2012). In this study, three classes of NBS genes in angiosperms also differed in intron characteristics (Fig. 4). Three introns with conserved positions and phases were shared by most TNL genes, with the first intron separating the TIR domain and the NBS domain (phase 2), the second intron separating the NBS domain and the whole LRR domain (phase 0), and the third intron separating the first and the remaining LRR domains (phase 0). Angiosperm RNLs share four class-specific introns, with the exception of RNLs in monocots, which did not possess the second intron (Fig. 4). Because the second intron is present in both A. trichopoda and dicot RNL genes, intron loss likely occurred in the monocot lineage. For the CNL class, conserved introns were not detected across all CNL lineages, especially in the basal lineages. In fact, a number of CNL genes were intronless, and were not concentrated in a specific lineage but distributed across all ancestral angiosperm nTNL lineages. Although some CNL genes had introns, these were more likely to have been gained at later evolutionary stages. This finding is in agreement with the intronless structure reported in bryophytes (Xue et al., 2012).

**The Three NBS Classes Show Distinct Evolutionary Patterns in Angiosperms**

Our step-by-step phylogenetic analysis not only supports the presence of three distinct NBS classes in the common ancestor of angiosperms, but also unraveled the ancient NBS gene lineages at multiple angiosperm evolutionary nodes (Fig. 5). Due to the loss of TNL genes in S. indicum and the monocot lineage, the TNL ancestral states at the asterid and mesangiospermae nodes could not be directly obtained from phylogenetic analyses. However, based on the recovered ancestral NBS lineages at the divergence nodes of Solanaceae (Supplemental Fig. S2), dicots (Supplemental Fig. S6), and all angiosperms (Supplemental Fig. S10), both asterid and mesangiospermae nodes were deduced to have at least five TNL lineages.

Above the family divergence nodes (An, Ma, Mo, Ed, As, Ro, and Er), CNL genes were abundant but only slightly more than those of earlier ancestral nodes (Fig. 5). This finding indicates that CNL genes adopted a gradual expansion strategy in the early stage of angiosperm evolution (from ~225 MYA as angiosperms diverged to around 100 MYA as asterids, rosids, and monocots diverged; Fig. 1; D’Hont et al., 2012; Wang et al., 2014; Zeng et al., 2014). Intensive expansion of CNL genes seems to have occurred in the familial ancestors of Solanaceae and Poaceae, reaching hundreds of copies. TNL genes apparently followed a different evolutionary path, as these were completely lost in monocots and in S. indicum. Although inherited by dicots, TNL numbers remained relatively constant for a long evolutionary period, with similar numbers of ancestral TNL genes identified at various ancestral nodes (An, Ma, Ed, As, Ro, and Er; Fig. 5). The expansion of TNLs mainly occurred prior to the divergence of families, such as in the Fabaceae ancestor, 55 TNLs were expanded from six of eight Ro-TNLs. A similar pattern was observed in the Brassicaceae ancestor, in which...
four Ro-TNL lineages expanding to 148 genes. The TNL genes were probably maintained at a lower number for more than 100 million years (from ~225 MYA as angiosperm diverged to around 100 MYA as rosids and asterids diverged) and only began to expand thereafter. This may explain why TNL genes were easily lost in multiple independent angiosperm lineages.

RNL genes are differentiated from the other two classes by their relatively conserved evolutionary pattern. Nearly all species from basal angiosperm, monocots, and asterids have one or two RNL genes, whereas a slight expansion of RNL genes was observed in certain rosid genomes (Fig. 1). Tracing the evolutionary history of RNL genes revealed two anciently diverged lineages in the common ancestor of angiosperms (Fig. 6). One RNL lineage contained the Arabidopsis ADR1 gene, whereas the other included the tobacco (Nicotiana benthamiana) NRG1 orthologs. Both lineages were inherited by dicot species, except for S. indicum, which lost the NRG1 lineage, similar to all monocot genomes. Several whole-genome duplications (WGDs) occurred throughout angiosperm evolution, causing gene duplications and functional divergence (Jiao et al., 2011; Vanneste et al., 2014). The separation of the ADR1 and NRG1 lineages prior to the divergence of angiosperms prompted us to investigate the involvement of an ancient WGD event. The syntenic relationships between chromosome fragments containing ADR1 or NRG1 orthologs were conserved in angiosperm genomes. Figure 6 (detailed in Supplemental Table S2) shows the synteny between ADR1- and NRG1-containing chromosomal regions between the dicot genomes of Phaseolus vulgaris (rosid lineage) and S. lycopersicum (asterid lineage). The syntenic blocks were also detected in the monocot genome of S. bicolor, in the absence of the NRG1 gene. Moreover, most RNL flanking gene pairs conserved in the duplicated blocks within the three species have high Ks values (>1; Supplemental Table S2), indicating that these genes may have resulted from one ancient WGD event. Because the most recent WGD event shared by monocots and dicots occurred in the common ancestor of all angiosperms (Jiao et al., 2011; Vanneste et al., 2014), the syntenic data and phylogenetic results support that RNL genes separated into two lineages sometime before this time point.

Deciphering the Evolution of Arabidopsis NBS Genes

By performing phylogenetic analyses at different evolutionary nodes, we have established a framework for the evolution of angiosperm NBS genes, which could help decipher the evolution of NBS genes in specific species as well as gain a much deeper view of the evolution of functionally characterized NBS genes.
In total, five out of 14 An-CNL genes were inherited by the Arabidopsis genome (Fig. 7). Gradual duplications occurred in four of these An-CNLs (An-C2, -C5, -C6, and C8), with An-C5 and An-C6 expanding to a few copies. Two An-CNLs (An-C2 and An-C8) produced more than 80% of Arabidopsis CNL genes, mainly occurring prior to the divergence of the Brassicaceae family. One CNL gene (At3g07040) represented an ancient gene (An-C14) that was conserved from the angiosperm ancestor to Arabidopsis without duplication. Together, the evolutionary paths of the five ancestral CNL genes resulted in the current CNL profile of Arabidopsis genome (Fig. 7).

The structural changes were also traced using the evolutionary framework of angiosperm NBS genes. For example, nearly all Arabidopsis CNL genes showed an intronless structure regardless of their origin (Meyers et al., 2003), which is in accordance with our proposed intronless structure for the ancestral angiosperm CNL gene. However, a group of Arabidopsis CNL genes were determined to possess two shared introns. All these genes originated from a lineage that diverged from An-C8 (Fig. 7) in the common ancestor of mesangiospermae, with the sister lineage remaining intronless. Based on the phylogenies of CNLs at different taxonomic levels, the two introns were frequently detected in orthologs of Brassicaceae species. The first intron also was found in orthologous genes of the legume family, grape, and the asterid species, but they were absent in monocots and A. trichopoda. Therefore, the exon-intron structure detected in these Arabidopsis CNL genes gradually formed from two independent intron gain events, one in the common ancestor of eudicots and the other acquired in the common ancestor of Brassicaceae (Fig. 7).

Furthermore, the angiosperm evolutionary framework of NBS genes enabled us to determine when a functional NBS gene likely arose during angiosperm evolution. Five functional CNL genes against different pathogens have been identified in Arabidopsis (Liu et al., 2007). According to the recovered evolutionary history of Arabidopsis CNL genes in Figure 7, RPM1 was the most ancient one among the five genes, possibly originating from the common ancestor of angiosperms, whereas the RPP8/HRT was most recent, possibly originating after Arabidopsis speciation. The origin of the other three genes could be traced to the divergence nodes of rosids, Brassicaceae, and Arabidopsis. Among the five genes, RPS2, RPS5, and RPM1 were R genes against P. syringae, whereas RPP8/HRT and RPP13 were both against Peronospora parasitica, indicating functional R genes continuously arose...
Figure 6. Evolutionary history of RNL genes inferred from phylogenetic and syntenic analyses. Top, The inheritance of the two RNL lineages in different angiosperm lineages. The inferred duplication is marked with a green circle, whereas inferred losses are indicated by black circles. Bottom, Intra- and intergenome syntenic blocks containing RNL genes detected within and among P. vulgaris, S. lycopersicum, and S. bicolor genomes.
against the same pathogen during plant evolution. Actually, apart from these genes, many other functional genes may have appeared in the history and then been lost during plant evolution. However, such cases often could not be traced due to their absence in model plant genomes.

The adoption of the angiosperm NBS evolutionary framework in Arabidopsis provides new insights into the evolutionary history of the NBS gene in this species. To facilitate the use of this framework in the analysis of NBS genes in other plants that were not included in this study, we generated hidden Markov model (HMM) profiles (Supplemental Dataset S1) for the 23 ancestral NBS lineages of angiosperms. Taken together, the angiosperm NBS evolutionary framework established in this study provides a fundamental reference for the

Figure 7. A few ancestral angiosperm NBS genes that are responsible in generating the current Arabidopsis CNL repertoire via gradual expansion. The ancient states of Arabidopsis at different evolutionary nodes were retrieved from the constructed NBS phylogenies. Offspring generated from the same ancestral gene at different divergence nodes are indicated by boxes of the same color. Black triangles indicate the birth of known functional genes. Gray arrows indicate intron gain events. The exon-intron structure of each ancestral lineage is shown.
Differential Expansion during Angiosperm Evolution

Although both TNL and CNL genes could be dated back to at least the arrival of land plants (Akita and Valkonen, 2002; Xue et al., 2012) and were inherited by the common ancestor of angiosperms, the TNL genes are apparently outnumbered by CNLs in most angiosperms, except for the Brassicaceae (Zhang et al., 2016). The phylogeny of TNLs also showed a lower level of topological complexity and a relatively shorter branch length compared to that in CNLs. All these discrepancies reflect different evolutionary patterns for TNLs and CNLs, and the results of this study suggest that the long-term contraction of TNLs and gradual expansion of CNLs in the early stages of angiosperm evolution (from ~225 MYA) may have been responsible for these differences. While <10 TNL genes were detected in rosids, eudicots, and probably mesangiospermae and asterid nodes, the number of CNL genes rapidly expanded to 35 early in the common ancestor of mesangiospermae and increased to 47 at the emergence of dicots (Fig. 5).

The long-term steady number of TNL genes also explains their independent loss in several angiosperm lineages. The absence of TNLs was first reported in cereals and major monocot lineages (Bai et al., 2002; Zhou et al., 2004; Yang et al., 2006; Li et al., 2010), then in two eudicot species, including an asterid, *Mimulus guttatus*, and a basal eudicot, *Aquilegia coerulescens* (Collier et al., 2011). Because both the basal angiosperm *A. trichopoda* and eudicots harbor TNL genes, it is highly likely that the monocot lineage lost its TNLs near the dicot/monocot split. *M. guttatus* and *S. indicum* are from two different families of Lamiales, indicating that the Lamiales lineage might have lost its TNLs soon after its separation with its sister lineage, Solanales. The presence of TNL genes in basal angiosperms and in both the rosid and asterid lineages of eudicots suggests that the loss of TNLs in the basal eudicot *A. coerulescens* also stems from an independent loss event. Therefore, at least three independent gene loss events occurred during angiosperm evolution. Furthermore, the near singleton status of TNL genes at the early stages of angiosperm evolution suggests that these events may have possibly occurred. Otherwise, deletion of an entire gene family consisting of dozens of members in three independent lineages would be unlikely, especially when these were dispersed in the genome.

Another interesting finding is that both CNLs and TNLs exhibit a recent expansion prior to the divergence of angiosperm families, with more than 100 NBS genes independently arising prior to the separation of each of the four families investigated in this study. Because the divergence of these angiosperm families occurred in the Paleogene and the divergence of their higher level taxonomies all occurred in the Cretaceous (Fig. 1; Lavin et al., 2005; Stefanovic et al., 2009; D’Hont et al., 2012; Zhang et al., 2012; Peng et al., 2013; Yang et al., 2013; Kim et al., 2014; Wang et al., 2014; Zeng et al., 2014), the observed intensive expansion of NBS genes in angiosperms more likely coincided with the Cretaceous-Paleogene (K-P) boundary (~66 MYA). This finding suggests that different lineages of angiosperms either underwent a process to reinforce resistance or experienced increased selective pressure from environmental pathogens during this period. The K-P boundary is an important period when mass extinction greatly impacted the evolution of various species. Angiosperms underwent drastic genomic and phenotypic changes to survive the climatic and environmental changes in this period, including prevalently detected WGDs across different angiosperm lineages and increased leaf vein density (Blonder et al., 2014; Vanneste et al., 2014). A hypothesis has been proposed that the documented bloom in fungal diversity during this period triggered the decline of reptiles, as well as promoted the expansion
of mammalian species, which have a more evolved immune system and body temperature (Vajda and McLoughlin, 2004; Casadevall, 2012). Actually, most fungal families diverged during the Cretaceous, and most fungal genera diverged at the K-P boundary, such as Botryosphaeriaceae and Fusarium, which are two fungal groups with many species that cause severe pathogenicity (Ma et al., 2013; Slippers et al., 2013; Hedges et al., 2015). Therefore, the fungal bloom during the K-P boundary may have enhanced the selection pressure on plant $R$ genes, which consequently accelerated NBS gene expansions at this period, prior to the divergence of angiosperm families.

An Ancient WGD Gave Rise to Two Functionally Divergent RNL Lineages

WGDs frequently occur during plant evolution (Jiao et al., 2011; Li et al., 2015; Qian et al., 2015). However, evidence from various studies has revealed that most NBS duplicons that originate from polyploidy are preferentially lost during evolution, possibly to avoid fitness costs (Bergelson et al., 2001; Cannon et al., 2004; Nobuta et al., 2005; Ashfield et al., 2012; Shao et al., 2014). Therefore, NBS genes resulting from ancient WGD are rarely detected. Our phylogenetic analysis has revealed that RNL genes were derived from two ancient lineages, namely, $ADR1$ and $NRG1$, which separated prior to the divergence of angiosperms. By surveying the syntenic regions in P. vulgaris and S. lycopersicum genomes, we determined that $ADR1$ and $NRG1$ orthologous genes were located on syntenic blocks in both genomes, suggesting that the two genes resulted from an ancient WGD. Despite having lost the $NRG1$ orthologs, the two syntenic blocks were also detected in the monocot genome of S. bicolor, suggesting that the WGD event was shared by dicots and monocots. Recent studies involving angiosperms have indicated that the most recent WGD shared by monocots and dicots occurred in the common ancestor of angiosperms (Jiao et al., 2011; Vanneste et al., 2014). Therefore, NBS genes resulting from ancient WGD are rarely detected. Our phylogenetic analysis has revealed that RNL genes were derived from two ancient lineages, namely, $ADR1$ and $NRG1$, which separated prior to the divergence of angiosperms. By surveying the syntenic regions in P. vulgaris and S. lycopersicum genomes, we determined that $ADR1$ and $NRG1$ orthologous genes were located on syntenic blocks in both genomes, suggesting that the two genes resulted from an ancient WGD. Despite having lost the $NRG1$ orthologs, the two syntenic blocks were also detected in the monocot genome of S. bicolor, suggesting that the WGD event was shared by dicots and monocots. Recent studies involving angiosperms have indicated that the most recent WGD shared by monocots and dicots occurred in the common ancestor of angiosperms (Jiao et al., 2011; Vanneste et al., 2014). Therefore, although syntenic blocks that contain $ADR1$ and $NRG1$ orthologs were not detected in A. trichopoda, our data suggest that the WGD contributing to the generation of two RNL lineages occurred in the ancestral angiosperm. This finding also is validated by the results of phylogenetic analysis, which places the one RNL gene in A. trichopoda in the $NRG1$ lineage with high statistical support but not basal to all RNL genes (Supplemental Fig. S9). The absence of an $ADR1$ orthologous gene in A. trichopoda might be attributable to either the incomplete coverage of the genome sequence or a species-specific gene loss. By searching the transcriptome data of the 1,000 plants project (1KP; www.onknp.com), we determined that both $ADR1$ and $NRG1$ orthologous genes are prevalently distributed in various basal angiosperms (data not shown). Taken together, both the phylogenetic analysis and the synteny results indicate that an ancient WGD involving the ancestral angiosperm gave rise to the two RNL lineages.

Functional RNL genes have been identified in both RNL lineages: $ADR1$ in Arabidopsis and $NRG1$ in tobacco. However, interestingly, neither $ADR1$ nor $NRG1$ functions as a canonical $R$ gene that detects specific pathogens (Collier et al., 2011). $ADR1$ and its paralogs are considered helper genes that transduce signals downstream from other $R$ genes (Bonardi et al., 2011). Similarly, NRG1 assists the $N$ gene in fighting against the Tobacco mosaic virus (Peart et al., 2005). Although neither $ADR1$ nor $NRG1$ acts as a pathogen detector, these are indispensable in resistance reactions. The distinct function, extremely conserved sequence, and gene structure suggest a downstream and crucial role for RNLs in disease resistance pathways. The simultaneous loss of TNL and $NRG1$ orthologs in M. guttatus, A. coerulea, and the Poaceae family has been speculated to be due to their close functional interdependence (Collier et al., 2011). It has been hypothesized that TNL genes only transfer their signal to $NRG1$ genes, whereas CNL genes could transfer their signal through either the $ADR1$ or $NRG1$ gene. The observed functional differentiation of $ADR1$ and $NRG1$ lineages suggests that divergence occurred after the two RNL copies emerged in the ancestral angiosperm WGD. As signal transduction protein-encoding genes, RNL genes are not involved in recognizing specific pathogens and thus are not under the selection of pathogens. Our results show that RNL lineages only kept a few copies in their genomes, except in several rosid species. The different evolutionary patterns among the three NBS classes reflect their distinct functions as either pathogen detectors or signal transducers, which further suggests that the NBS family diverged early to contribute to the plant immune system from multiple aspects.

An Angiosperm Framework for Understanding NBS Gene Evolution

Most previous studies of NBS genes were mainly performed on individual species, which hindered in better understanding NBS gene evolution at the whole-angiosperm scale. The results reported in this study (covering $>6,000$ NBS genes from 22 genomes) provide a fundamental phylogenetic framework of NBS genes in angiosperms. Because the whole dataset is too large for phylogenetic analysis and subsequent gene loss/gain reconcile, we used a step-by-step method for phylogenetic analysis as described earlier. In such procedure, the phylogenetic relationships among all angiosperm NBS genes could not be fully displayed in a single tree, and supporting values for some internal nodes may decrease because only a few representative sequences were used in the construction of high taxonomic level NBS phylogenies. However, our data suggest that the results from this approach are generally robust, and the taxonomy of NBS genes often were not affected by different datasets used. For example, one of our major findings, the sister relationship of CNL and RNL, is well supported in nTNL phylogenies constructed by different datasets (Supplemental Figs. S1, S2).
Evolution of NBS-LRR Genes in Angiosperms

S3, S4, S6, and S9). Furthermore, during the procedure, with more data incorporated, the shared introns in a CNL lineage (An-C8; Fig. 7) also were in perfect agreement with its phylogenetic relationships. Therefore, to a large extent, the phylogenetic framework of angiosperm NBS provided in this study is reliable for the estimation of the evolutionary history of this large gene family. Using this framework as reference, NBS gene profiles in a particular plant genome could be fully classified to better understand the evolutionary dynamics. For example, the 52 CNL genes present in the Arabidopsis genome were traced to five ancestral CNL genes in the common ancestor of angiosperms (Fig. 7). Gradual changes in gene content in these ancestral lineages were observed at five intermediate divergence nodes and the node to Arabidopsis speciation, which not only revealed different fates for the five ancestral genes during their evolution but also uncovered the extent of gene expansion at different nodes. Furthermore, the five known Arabidopsis functional CNL genes were traced to four ancient NBS genes and arose from different angiosperm nodes (Fig. 7). Notably, two of the five R genes, RPS2 and RPM1, defend against *P. syringae* by guarding the Arabidopsis RIN4 protein. The state change of RIN4 that is monitored by RPS2 and RPM1 also are modified by two different effectors of *P. syringae* (Mackey et al., 2002; Mackey et al., 2003), suggesting a long-term coevolutionary arms race among Arabiopsis R genes and *P. syringae* (Fig. 7). Angiosperms may have first evolved the RPM1 gene as a mechanism against *P. syringae* infection by detecting effector (AvrRpm1)-induced phosphorylation of RIN4. On the other hand, *P. syringae* evolved another effector (AvrRpt2) to cleave the RIN4 protein and evade the activation of RPM1, and plants further evolved another CNL gene (RPS2, probably prior to the node of rosid divergence) to monitor the cleavage of RIN4 and activate the immune system. The long history of R genes guarding RIN4 also suggests that this protein is an ancient target of pathogens, which in turn unravels its important function in plant immune signal transduction (Liu et al., 2009; Hou et al., 2011). It should be noticed that although we traced the emergence time of these genes in angiosperm lineages, we were not able to accurately determine when they obtained specific functions. The resistance specificity may have existed prior to the duplication event but subsequently been lost from some duplicated copy; alternatively, the specificity may have arisen after the duplication event.

In conclusion, by identifying more than 6,000 NBS genes from 22 angiosperm species, this study has successfully constructed the phylogeny of angiosperm NBS genes. A total of 25 ancestral NBS genes were identified in angiosperm ancestors, and three distinct NBS classes were revealed. Although exhibiting different evolutionary patterns, both TNL and CNL genes underwent extensive expansion at the K-P boundary. In contrast, RNL genes conservatively evolved during angiosperm evolution, which mirrors their distinct pathogen detection or signal transduction roles in plant disease resistance. The angiosperm level NBS phylogenetic framework constructed in this study provides a fundamental reference for further studies on NBS genes.

**MATERIALS AND METHODS**

**Database**

The genomic sequences, annotations, and gene models for 13 out of 22 genomes (except four Fabaceae and five Brassicaceae genomes) were used in this study. Among these, data on *Vitis vinifera*, *Solanium lycopersicum*, *Solannum tuberosum*, *Brachypodium distachyon*, *Setaria italica*, *Sorghum bicolor*, *Oryza sativa*, and *Zea mays* were downloaded from the Phytozome database (http://www.phytozome.net/; Goodstein et al., 2012). Files of *Amborella trichopoda*, *Capsicum annuum*, *Sesamum indicum*, *Phyllostachys heterocycla*, and *Musa acuminate* were downloaded from the Pepper Genome Database (http://peppersequence.genomics.com.cn/page/species/index.jsp; Qin et al., 2014), Sibbase (http://ori-genomics.org/Sinbase/manual.htm; Wang et al., 2014), BambooDB (http://www.bamboogdb.org/page/download.jsp; Peng et al., 2013), Banana Genome Hub (http://banana-genome.cirad.fr; D’Hont et al., 2012), and the Amborella genome database (http://www.amborella.org; Amborella-Genome-Project, 2013), respectively.

**Identification of NBS-Encoding Genes**

The NBS genes in the 13 genomes were identified as described in previous studies (Shao et al., 2013; Shao et al., 2015; Zhang et al., 2016). Briefly, a HMM search was first performed for protein sequences in each genome using hmmer3.0 (http://hmmmer.org) using default parameter settings, with the NB-ARC domain (Pfam: PF00931) as query. The amino acid sequence of the NB-ARC domain was then used to run a BLASTp search against all protein sequences in each genome, with the threshold expectation value set to 1.0. All hits obtained using HMM or BLAST searches were then merged together, and the redundant hits were removed. The remaining sequences were further subjected to Pfam analysis (http://pfam.sanger.ac.uk/) to further exclude sequences that do not have a detectable NBS domain at an E value of 10~^3~. When two or more transcripts were annotated for a gene from alternative splicing, the longest form with an NBS domain was selected. Finally, the identified NBS domain-encoding genes were examined to see whether these encode the TIR, RWP8, CC, or LRR domains using Pfam and COILS analyses (Lupas et al., 1991). The exon position and intron phase of each gene were transformed from the gff3 file of the reference genome.

**Sequence Alignment and Phylogenetic Analysis**

Sequence alignment and phylogenetic analysis were performed as described previously (Shao et al., 2014). Amino acid sequences of the NBS domain were aligned using ClustalW (Edgar, 2004) with default options (Thompson et al., 2012) and then manually corrected in MEGA 5.0 (Tamura et al., 2011). The resulting amino acid sequence alignments were used to guide the alignments of the nucleotide coding sequences. Genes with very short NBS domains were eliminated from the matrix because these interfered in line analysis and subsequent construction of phylogenetic trees. All alignments used for phylogenetic analysis in the presented study are provided in Supplemental Dataset S2. Phylogenetic analysis was conducted using a maximum likelihood method as provided in the Phylm program (Guindon and Gascuel, 2003), which was integrated in the Seaview package (Gouy et al., 2010). Branch support was assessed with the aLRT statistic (Guindon et al., 2010). Gene losses and gains at each divergence node were reconciled using Notung software (Stoltzer et al., 2012). The number of ancestral NBS genes at different divergence nodes was determined using monophyletic NBS lineages as described elsewhere (Shao et al., 2013).

Because merging all data generates large datasets for both nTNL and TNL classes, we adopted a step-by-step strategy to recover the full picture of NBS gene evolution in angiosperms. Generally, phylogenetic analysis was first performed at a low taxonomy level, mainly containing NBS genes from a few genomes. Then, representative NBS sequences from these angiosperm lineages were compiled with NBS genes in more plant genomes to construct the NBS gene phylogenies at higher taxonomy levels, finally reaching the whole-angiosperm level.
Supplemental Data

The following supplemental materials are available.

Supplemental Figure S1. Phylogenetic tree of monocot nTNL genes.

Supplemental Figure S2. Phylogenetic tree of asterid nTNL genes.

Supplemental Figure S3. Phylogenetic tree of asterid TNL genes.

Supplemental Figure S4. Phylogenetic tree of rosid nTNL genes.

Supplemental Figure S5. Phylogenetic tree of rosid TNL genes.

Supplemental Figure S6. Phylogenetic tree of dicot nTNL genes.

Supplemental Figure S7. Phylogenetic tree of dicot TNL genes.

Supplemental Figure S8. Phylogenetic tree of mesangiospermae nTNL genes.

Supplemental Figure S9. Phylogenetic tree of angiosperm nTNL genes.

Supplemental Figure S10. Phylogenetic tree of angiosperm TNL genes.

Supplemental Table S1. NBS-encoding genes detected in angiosperm genomes examined in this study.

Supplemental Table S2. Inter- or intraspecies RNL-containing syntenic blocks.

Supplemental Dataset S1. HMM profiles for the 23 ancestral NBS lineages.

Supplemental Dataset S2. Alignments used for phylogenetic analysis in this study.

Received September 18, 2015; accepted February 1, 2016; published February 2, 2016.

LITERATURE CITED


Downloaded on May 27, 2021. - Published by plantphysiol.org
Copyright (c) 2020 American Society of Plant Biologists. All rights reserved.


