Systematic Analysis of NB-ARC Gene Family in Rice and Functional Characterization of *GNP12*

Ying-Hua Pan (✉ panyinghua2008@163.com)

Guangxi Academy of Agricultural Sciences

Lei Chen

Guangxi Academy of Agricultural Sciences

Hai-Feng Guo

China Agricultural University

Qi-Jin Lou

China Agricultural University

Rui Feng

Guangxi Academy of Agricultural Sciences

Muhammad Abdul Rehman Rashid

Government College University Faisalabad

Xiao-Yang Zhu

China Agricultural University

Dong-Jin Qing

Guangxi Academy of Agricultural Sciences

Hai-Fu Liang

Guangxi Academy of Agricultural Sciences

Li-Jun Gao

Guangxi Academy of Agricultural Sciences

Cheng-Cui Huang

Guangxi Academy of Agricultural Sciences

Yan Zhao

Shandong Agricultural University

Guo-Fu Deng

Guangxi Academy of Agricultural Sciences

Research Article

Keywords: NB-ARC family, Evolution, Expression, Grain number of panicle, Rice grain length

Posted Date: February 23rd, 2022

DOI: <https://doi.org/10.21203/rs.3.rs-1379454/v1>

License: This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Abstract

The NB-ARC (nucleotide-binding adaptor shared by APAF-1, R proteins, and CED-4) gene family plays a critical role in plant development. However, our understanding of NB-ARC genes regulation of plant development mechanisms in the plant panicle is still limited. Here, we subjected 258 NB-ARC genes in rice to genome-wide analysis to characterize their structure, function, and expression patterns. The NB-ARC genes were classified into three major groups, and group II included nine subgroups. Evolutionary analysis of NB-ARC genes in dicotyledon plant (*Arabidopsis thaliana*) and monocotyledonous plants (*Oryza sativa* L. and *Triticum aestivum*) indicate that orthologous genome segments were conserved in monocotyledons and subjected to weak positive selective pressure during evolution. Dispersed and proximal replication events were detected. Expression analysis showed expression of most NB-ARC genes in roots, panicles, and leaves, and regulation at the panicle development stage in rice Ce253. The *GNP12* gene encodes RGH1A protein, which regulates rice yield according to panicle length, grain number of panicle, and grain length, with eight major haplotypes. Most NB-ARC protein family members are predicted to contain P-loop conserved domains and are located in the membrane. The results of this study provided insight into the characteristics and evolution of NB-ARC family and suggest that *GNP12* positively regulates panicle development.

Background

NB-ARC (nucleotide-binding adaptor shared by APAF-1, R proteins, and CED-4) proteins are critical regulators of signaling pathways and play important roles in effector recognition and signal transduction in healthy plant growth and development (Chen et al. 2018; Jacquemin et al. 2014; Wen et al. 2017). NB-ARC proteins contain an NB-ARC domain that is proposed to act as a molecular switch, with a nucleotide-binding (NB) subdomain including a P-loop, a C-terminal extension that forms a four-helix bundle (ARC1), and a winged-helix fold (Albrecht et al. 2006; van der Biezen et al. 1998). The NB subdomain can bind and hydrolyze ATP *in vitro* (Tameling, 2002). NB-ARC belongs to the STAND (signal transduction ATPases with numerous domains) family of NTPases, and is proposed to work as an NTP-hydrolyzing switch by binding and hydrolyzing NTP and regulating signal transduction by conformational changes (Leipe et al. 2004). Recent studies revealed NB-ARC proteins with similar central nucleotide-binding-ARC domain architectures act in metazoan innate immunity and programmed cell death pathways (Slootweg et al. 2009). For example, NB-ARC genes share homology with human APAF-1 and *C. elegans* CED-4, proteins that regulate cell death (Yildirim-Ersoy et al. 2011). RPP1A belongs to the NB-ARC gene family and reduces plant growth with broad-spectrum resistance to virulent strains of *H. parasitica* in *Arabidopsis thaliana* (Michael et al. 2006). Constitutive expression of TIR-NB-ARC-LRR gene *VpTNL1* in *Arabidopsis thaliana* resulted in either a wild-type or dwarf phenotype (Wen et al. 2017). NB-ARC genes might contribute to *R. glutinosa* consecutive monoculture problems (Chen et al. 2018). Domains can work together as a platform to mediate downstream signal transduction events (Lukasik et al. 2009). In wheat, many NB-ARC gene clusters include highly similar genes likely formed by tandem duplications (Andersen et al. 2020). NB-ARC-containing sequences in wheat exhibits maximum homology with the proteins from

Oryza sativa indica and *Brachypodium distachyon*, and distributions of NB-ARC sequences are balanced among the three wheat sub-genomes (Chandra et al. 2017). In *Oryza sativa* (Erik Sloodweg), *Oryza glaberrima* (Sloodweg et al. 2009), and *Oryza brachyantha* (FF), conserved NB-ARC genes were subjected to strong purifying selection but selection was more relaxed for expanded orthologous genes (Jacquemin et al. 2014). NB-ARC protein function has been studied in wheat, rice, and *Arabidopsis thaliana*, but the mechanism of NB-ARC protein function in rice panicle development remains poorly understood.

Some NB genes have been reported that regulated in response to morphological development (IgariK et al. 2008). *AhRAF4* was also observed to correlate with morphological changes in development in maize (Dolezal et al. 2014). Rice (*Oryza sativa* L.) is the world's most important cereal crop and is consumed by half of the world's population (Zhao et al. 2015). In recent years, many important genes controlling rice grain yield have been isolated and functionally characterized. *LP1* encodes a Remorin_C-containing protein of unknown function and the *LP1* allele of Xiushui79 leads to reduced panicle length (Liu et al. 2016). Grain Number per Panicle1 (*GNP1*), Rice *GA20ox1* encoding a cytokinin (Tameling et al. 2002), and the biosynthesis gene for gibberellins upregulate cytokinin activity to increase grain number and grain yield in rice (Wu et al. 2016). *GW8* encodes an SBP-domain transcription factor that regulates grain width by binding directly to the *GW7* promoter to repress its expression (Wang et al. 2015). *OsMKK3* encodes a MAP kinase that controls rice grain size and chalkiness (Pan et al. 2021). Although few studies have focused on NB-ARC proteins in rice, these important plant proteins likely have multiple regulatory roles. For example, the gain-of-function mutation of NB-ARC protein *RLS1* (*Rapid Leaf Senescence1*) causes high-light-dependent HR-like cell death in rice (Wang et al. 2020). Overall, a greater understanding of mechanisms underlying grain yield will strengthen our understanding of regulatory mechanisms for these traits and facilitate the breeding of crop varieties with high-yield potential.

In this study, members of the NB-ARC gene family were identified from analysis of the rice genome. The motif composition and gene structures of these genes were systematically analyzed, as well as tandem duplication and gene replication events. Collinear relationships between *Oryza sativa*, *Triticum aestivum*, and *Arabidopsis thaliana* were compared. The expression levels of 11 subgroups of NB-ARC genes in different tissues and panicle development of rice Ce253 were analyzed by RNA-Seq and the expression levels of 18 genes were measured via qPCR. Functional analysis was performed of a selected NB-ARC gene, *GNP12*. This gene serves as a positive regulator in panicle development and panicle length. The results of this research provide a better understanding of the function of the NB-ARC gene family and provide guidance for future efforts to improve rice breeding.

Results

NB-ARC genes in rice were analyzed and subgroup II_f was identified as a representative subgroup

Hidden Markov Model and a Simple Modular Architecture Research Tool search with Pfam database (PF00931.2) were used to identify a total of 258 rice NB-ARC genes in the *Oryza sativa* genome

(Additional file 1: Table S1). The lengths of the CDS for these genes range from 624 (*LOC_Os02g27680*) to 7773 (*LOC_Os06g41690*) bp. Analysis revealed that 49.22% of these NB-ARC genes contain 2–6 introns and 86.04% NB-ARC genes contain 1–4 exons (Additional file 1: Table S1). Most of the NB-ARC genes have a long CDS and show a highly conserved structure (Fig. 1). To explore the expansion of NB-ARC family members in rice, these sequences were used to generate an unrooted phylogenetic tree. The NB-ARC family members were divided into three major groups. Group I contained 29 genes, Group II contained 226 genes, and group III contained only three genes. The NB-ARC genes in group II were further classified into nine subgroups (groups IIa, IIb, IIc, IId, IIe, IIlf, IIlg, IIh, and Ili). Of these, the two groups with the most members were IIlf with 35 genes and Ili with 37 genes. Group IIc contained the fewest number of genes, 13. A total of 15 motifs were identified by MEME, and two of these, M2 and M6, were present in 98.83% and 97.28%, respectively, of the NB-ARC genes. Genes on a single branch contained similar numbers of introns and similar distribution and number of motifs. In a given subtribe of genes, the positions of motifs were highly conserved. In group IIlf, the genes contained 1–4 introns and 2–12 exons. Most genes in this group share 10 similar motifs (motifs 1, 2, 3, 4, 6, 8, 9, 10, 11, and 12) (Fig. 1). Group IIlf was representative in motifs and numbers of NB-ARC genes. Overall, NB-ARC genes were unevenly distributed across all 12 chromosomes, with 29 (11.24%), 61 (23.64%), and 34 (13.18%) genes located on Chr 8, 11, and 12, respectively. (Fig. 2a). Gene analysis revealed duplication events including segmental, tandem, proximal, and dispersed, but not singleton duplications. Dispersed (110) and proximal replication (82) most frequently occurred in NB-ARC genes across all chromosomes (Fig. 2b). Segmental replication was detected on Chr2, Chr4, Chr7, and Chr11 only (Fig. 2b). Pairwise replication of genes was also detected in NB-ARC genes (Fig. 2c). The genes in Group IIlf, mapped to eight chromosomes, with dispersed, tandem, and proximal replication. In general, subgroup IIlf of group II showed typical structure with main motif and contain more genes, and analysis suggested that replication events were the main driving force of NB-ARC evolution.

Conservation of NB-ARC orthologous genome segments in monocotyledons with weak positive selective pressure during evolution

To study the relationships between the NB-ARC genes of rice and other model plant species, including gramineous plants, we performed cluster analysis (Fig. 3a). The NB-ARC genes from, a dicotyledon plant (*A. thaliana*) and two monocotyledonous plants (*Oryza sativa* L. and *Triticum aestivum*) were compared. Pfam was used to compare 1052 and 50 NB-ARC genes in *Triticum aestivum* and *A. thaliana*, respectively. Next, based on the highly conserved NB-ARC domains of rice, *Triticum aestivum*, and *A. thaliana*, a phylogenetic tree was built. The 50 genes of *A. thaliana* belonged to a single subgroup, and every subgroup contained rice and *Triticum aestivum* genes (Fig. 3a). Within each subgroup, genes from the homologous chromosome group generally clustered into a clade (Fig. 3a); for example, *LocOs12g36690*, *LocOs12g36720*, and *LocOs12g33160* were clustered in a subgroup. To further study the relationships between the NB-ARC genes of rice and other plants, we next performed whole genome synteny analysis (Fig. 3b). A total of 65 orthologous genome segments were distributed on 12 chromosomes in *Triticum aestivum* (Additional file 2: Table S2). The highest number of orthologous genome segments, 19, mapped to the syntenic locus in chromosome 4 and *Triticum aestivum*

chromosomes 2A, 2B, 2D. Two orthologous genome segments mapped to chromosome 5 in *A. thaliana* (Fig. 3b) For further evolutionary analysis, the Ka, Ks, and Ka/Ks values of orthologous gene pairs were calculated based on the comparative synteny map (Fig. 3c). Ka/Ks ratios of 2571 gene pairs were evaluated in rice and *Triticum aestivum* for their evolution in protein coding genes. Fifty-two gene pairs had Ka/Ks > 1, 2412 gene pairs had 0.5 < Ka/Ks < 1, no gene pairs had < 0.1 Ka/Ks ratios were similarly evaluated for 70 gene pairs in rice and *A. thaliana* for their evolution in protein coding genes. Five gene pairs had Ka/Ks > 1, 59 gene pairs had 0.5 < Ka/Ks < 1, no gene pairs had < 0.1, and the remaining had no data. In rice, 15 gene pairs had Ka/Ks > 1, 431 gene pairs had 0.5 < Ka/Ks < 1, no gene pairs had < 0.1, and five gene pairs had 0.1 < Ka/Ks < 0.5 (Additional file 3: Table S3). The majority of orthologous NB-ARC gene pairs had 0.5 < Ka/Ks < 1, suggesting that the NB-ARC gene family in these three plant species experienced weak positive selective pressure during evolution. Evolution of the few NB-ARC genes in rice was governed by strong constraints that may have contributed to their structural and functional stability (Fig. 2c; Additional file 4: Table S4). Interestingly, some orthologous genome segments mapped between rice and *Triticum aestivum* were not observed between rice and *A. thaliana*, which may indicate that these orthologous pairs formed after the divergence of dicotyledonous and monocotyledonous plants. Overall, the analysis shows that the NB-ARC gene family is highly conserved in monocotyledons and orthologous genome segments suggest that segmental duplications may be the main cause of the extension of this gene family. Multiple copies of genes have arisen during the evolution of dicotyledons and monocotyledons.

Most Nb-arc Genes Regulate Panicle Development

To study the spatial and temporal expression patterns of NB-ARC genes, transcriptomic profiling was profiled across eight different tissues (anther, callus, leaf, panicle, pistil, root, seed, and shoot) to analyze the role of these genes in rice organ development according to the Rice Expression Database (Additional file 5: Table S5; Fig. 4a). In our findings, 23 genes exhibited no expression in any of the tissues, five genes showed expression in eight tissues, and the remaining 135 genes showed expressional activation in more than three tissues. Expression varied for different tissues, with expression of 143 genes in root, 94 genes in panicle, and only nine genes in anther. For the Ilf subgroup, 15 genes were most expressed in callus and 20 genes were most expressed in panicle. Twelve genes of Ilf and 12 genes of Ili were most expressed in pistil and 27 genes of Ilf were most expressed in root. Ilf, a typical subgroup in the NB-ARC gene family, includes genes such as *LOC_Os12g36720* that was expressed in eight different tissues (Fig. 4a). To confirm the expression patterns, expression of 11 genes of different subgroups were detected by qRT-PCR. The results showed that six of these genes exhibited similar expression patterns to the transcriptomic data (Fig. 4b). Most of these genes showed high expression in root. *LOC_Os12g36720* in Ilf and *LOC_Os06g22460* in Ile were most highly expressed in anther and panicle.

Few studies have explored the roles of NB-ARC genes in normal plant growth and development. To investigate a possible function of these genes to regulate plant panicle development, the expression of selected NB-ARC genes was analyzed at the panicle development stage of cultivated rice (Fig. 4b, c). In

rice Ce253, 7 genes were upregulated in 5, 10, 15, and 20 cm panicle length stages and downregulated in the 25 cm panicle length development stage. Twenty-five genes (four genes of group I, six genes of IIa, one gene of IIb, two genes of IIc, two genes of IId, two genes of IIe, one gene of IIf, four genes of IIg, two genes of IIh, and one gene of IIi) were downregulated throughout panicle development. Genes in group IIf clustered into three subgroups based on expression. Most genes of the first subgroup of IIf were down-regulated at the early stage of young panicle differentiation, up-regulated during rapid panicle lengthening and grain type elongation, and then down-regulated at the stage of stable panicle length (20–25 cm of panicle length) (Fig. 4b). This first IIf subgroup (39.39%) includes genes such as *LOC_Os12g36690*, *LOC_Os11g45970*, and *LOC_Os12g36720* and the observed expressional activation in panicle development for these genes suggests a crucial role in panicle development of rice.

GNP12 regulates rice yield with effects on panicle length, panicle grain number, and grain length

To test the ability of NB-ARC genes to regulate plant panicle development, *LOC_Os12g36720* as a representative IIf gene were identified for showing expression in anther; callus, leaf, panicle, pistil, root, seed, and shoot; and exhibited a remarkably high level of expression in Ce253 during panicle development. *LOC_Os12g36720* encodes a Resistance Gene Homologs (*RGH1A*) gene. Therefore, *LOC_Os12g36720*, *GNP12*, was selected for further characterization (Additional file 5: Table S5). We used a CRISPR-Cas9 system for targeted gene mutation of *GNP12* in Ce253 (Cr line) (Fig. 5a). The resulting deletions lead to frameshifting mutations that result in incomplete peptides of *GNP12*; these truncated peptides exhibit a loss of *NB-ARC* domain function. Compared to Ce253, the three Cr lines Cr-1, Cr-2, Cr-3, were reduced 13.68, 3.71, and 9.39% in panicle length (Fig. 5b); 23.80, 8.50, and 24% in panicle grain number; 2.76, 0.7, and 2.51% in grain length; and 26.58, 8.79, and 10.98% in setting percentage, respectively. The three Cr lines showed significant reductions in panicle length, grain number of panicle, grain length, and setting percentage (Fig. 5a, b, e, f, h, j). To investigate the tissue structures affected by *GNP12*, microstructures of grain chaff were observed by scanning electron microscopy (Fig. 5d). The chaff of Cr was shorter than that of Ce253 (Fig. 5d), but the Cr lines did not differ in length-width grain ratio or grain width (Fig. 5c, i, k).

Cytokinin (Tameling et al. 2002) is a primary determinant of plant architecture. To investigate whether *GNP12* influences CTK levels, we compared the CTK content in wild-type and Cr lines. The CTK content in Cr lines was reduced to 30.28% of that in the wild-type (Fig. 5l). The results suggest that *GNP12* may affect panicle development and grain production in rice through changes in CTK content.

To investigate natural variation in *GNP12*, we analyzed the polymorphism of *GNP12* in the 3K rice genomes (3K-RG) dataset (Wang et al. 2018). We identified eight major haplotypes of *GNP12* with 32 single-nucleotide polymorphisms (SNPs) in the coding sequence region in 827 accessions of the 3K-RG dataset (Fig. 5m; Table 1). Except for Hap2, the frequencies of the major haplotypes differed significantly among different subgroups. Hap1, Hap5, Hap6, and Hap7 were specifically found in *indica* rice cultivars, with Hap1 being the most prevalent. Hap2 was mainly identified in *indica*, temperate *japonica*, tropical *japonica*, and Bas rice cultivars. Hap3 was mainly present in tropical *japonica* and *indica* rice cultivars.

Hap4 and Hap8 were mainly found in Aus rice cultivars (Fig. 5n). Phenotypic evaluation showed that Hap1, Hap3, Hap5, and Hap6 in the *indica* subgroup and Hap3 in the *japonica* subgroup showed longer grain length than other haplotypes. Haplotype network analysis showed that favorable haplotypes of *GNP12* may have different origins. Favorable *indica* haplotypes Hap3 and Hap6 may have come directly from *O. rufipogon* I and favorable *indica* haplotypes Hap1 and Hap5 may have come from novel favorable mutations during *indica* domestication. Favorable haplotype Hap3 in *japonica* may have arisen during the differentiation of temperate *japonica* and tropical *japonica* (Fig. 5o).

To determine whether selection acted on *GNP12*, we analyzed the nucleotide diversity of *GNP12* among different rice subgroups. On average, the nucleotide diversities of *GNP12* in tropical *japonica* and *indica* subgroups were respectively much lower and higher than those in other haplotypes. To assess whether the DNA sequence evolved randomly or under a non-random process, Tajima's D values were calculated for *GNP12* in tropical *japonica* and *indica* and were negative and positive, respectively. Both values deviated significantly from zero, implying potential positive selection and balancing selection acting on *GNP12* in tropical *japonica* and *indica*, respectively (Fig. 5p). To further test whether the observed reduction in nucleotide diversity in tropical *japonica* was due to positive selection or a bottleneck effect, we calculated the nucleotide diversity of 40-kb flanking regions of *GNP12*. We found that the average nucleotide diversity of the *GNP12* flanking regions in tropical *japonica* was significantly lower than that in other regions, suggesting that the decrease of nucleotide diversity of *GNP12* in tropical *japonica* may be largely caused by positive selection (Table 2).

Nb-arc Proteins Contain Conserved P-loop Domains

To further investigate the subdomains and functions of NB-ARC genes, 11 predicted proteins from group I, group II, and group III were selected for further analysis. Strong conservation was identified for several subdomains: α C, α M, α S, β 3, β 2- β 5, and P-loop (Fig. 6a, b). The most conserved residues in the putative NB-ARC domains were found in the P-loop (G193 and S223) and α S (L468) subdomains (Fig. 6a). Analysis of the gene structure and conserved motifs revealed highly conservation of rice NB ARC genes with no obvious variation in the amino acid residues of the P-loop region among homologous genes (Fig. 6c). We observed few variants of the 11 amino acids in the P-loop core, with substantial variation in genic frequency of these variants in rice. These results indicate the significant conservation of the P-loop in most rice NB-ARC proteins. The predicted 3D structures of the identified rice NB-ARC proteins were conserved, which was consistent with the phylogenetic, gene structure, and conserved domain analysis. *LOC_Os02g02670*, *LOC_Os06g41670*, *LOC_Os06g22460*, *LOC_Os012g36720*, *LOC_Os12g17490.1*, *LOC_Os10g10360* were predicted to share similar structures (Fig. 6c). Predicted models were constructed to heuristically maximize alignment coverage, identity percentage, and confidence score for the tested sequences. For the 11 predicted NB-ARC proteins, α -helix was most highly predicted as secondary structure (37.41–63.56%), followed by random coil (20.03–35.64%), extended strands (7.3-20.84%) and β -strands (0-5.29%) (Fig. 6c, Additional file 6: Table S6). The 3D modeling results revealed similarity of predicted tertiary structures, implying that NB-ARC proteins may have evolved from a shared ancestor

and/or under purification selection force for stabilization during long-term acclimation after initial divergence.

To further characterize the function of the NB-ARC genes, subcellular localization studies were performed. We randomly selected three NB-ARC genes, *LOC_Os12g36720* (Ilf), *LOC_Os101g458510* (Ili), for subcellular localization in rice protoplast cells. *LOC_Os12g36720*-GFP, *LOC_Os101g458510*-GFP expression vectors were constructed and transformed into rice protoplasts for transient expression. Control GFP expression was observed throughout the membrane, cytoplasm and nucleus, the *GNP12* and *LOC_Os101g458510* proteins were mainly expressed in the nucleus and membrane respectively. The result indicating that these proteins are widely present in various organelles.

Discussion

Conserved domains and polymorphisms of NB-ARC genes maintained a dynamic balance during evolution

NB-ARC proteins play central roles in recognizing pathogens, initiating defense cascades, and maintaining cell development. The NB-ARC conserved domain maintains stable similar structures in different species, likely due to convergent evolution, suggesting these proteins may perform similar recognition and activation mechanisms (Slootweg et al. 2009). In the Fabaceae family, eight conserved motifs of NB-ARC domains have been identified (Pal et al. 2007). Plant NOD-like receptor (NLR) proteins contain NB-ARC domains with structural similarities to their mammalian homologues (Steele et al. 2019). However, there is high polymorphism in some NB-ARC genes in some species. The combination of conserved motifs and the evolution of rich polymorphisms may allow response to environmental signal stimulation to achieve coevolution by replication events. In *Arabidopsis*, analysis of the NLR family revealed that positive selection and recombination occurred frequently in the leucine-rich repeat (LRR) domain but there was negative selection in the nucleotide-binding (NB-ARC) domain (Mondragón-Palomino et al. 2017). In peanut, *AhRAF4* of NB-ARC proteins evolved by recombination with duplications and point mutations from *Arachis duranensis* (Deng et al. 2018). Among *O. sativa* (Slootweg et al. 2009), *O. glaberrima* (Slootweg et al. 2009), and *O. brachyantha* (FF), a high number of inparalogs suggests that the NB-ARC family experienced highly dynamic evolution, with a large number of tandem arrays and duplicated genes observed in the *O. sativa* subspecies (Jacquemin et al. 2014). NB-ARC genes diversified through duplication to encode receptors adapted to external signals (Andersen et al. 2020). IPm21 encodes a coiled-coil, nucleotide-binding site, leucine-rich repeat (CC-NBS-LRR) protein, and evolutionary analysis of 38 non-redundant Pm21 alleles indicated that the nucleotide diversity of the LRR domain was significantly higher than those of the CC and NB-ARC domains (He et al. 2020). NB-ARC genes can form genomic clusters adjacent to LRR-RLK-XILs by mechanisms other than genomic clustering (Ngou et al. 2022). Overall, the evolution of NB-ARC genes maintains key function through conservation of the NB-ARC domain while the C-terminal regions of these proteins exhibit polymorphisms to allow these genes to co-evolve with the environment through replication events.

GNP12 may regulate rice yield by influencing hormone activity

Panicle grain number and length of panicle are important agronomic characteristics, but the genetic determinants of these traits remain unclear. Many genes regulate grain number per panicle, length of panicle, and grain length by influencing hormone activity. *ERECTA1* (*OsER1*) is a cytokinin oxidase/dehydrogenase that negatively regulates grain number per panicle (Ashikari et al. 2005). *ERECTA1* acts upstream of the *OsMKKK10-OsMKK4-OsMPK6* cascade, and the *OsER1-OsMKKK10-OsMKK4-OsMPK6* pathway is required to maintain cytokinin homeostasis (Guo et al. 2020). *OsASP1* encodes a TOPLESS-related transcriptional co-repressor, is closely associated with auxin action, and regulates spikelet development (Yoshida et al. 2012). *TGW6* encodes a protein with indole-3-acetic acid (IAA)-glucose hydrolase activity that can control IAA supply at the transition from the syncytial to the cellular phase to limit cell number and grain length (Ishimaru et al. 2013). *GW6* encodes a GA-regulated GAST family protein and positively regulates grain width and weight (Shi et al. 2020).

NB-ARC gene regulates plant development by catalyzing ADP through ARC structure. NB-ARC proteins carry a central nucleotide-binding-ARC domain that binds ADP/ATP rather than GDP/GTP (Chattopadhyaya et al. 2008). The NB-ARC domain may be a molecular switch between ADP (repressed) and ATP (active) binding forms (Steele et al. 2019). A tomato (*Lycopersicon esculentum*) R protein I-2 with NB-ARC domain is impaired in ATP hydrolysis, but not in ATP binding, suggesting a molecular switch whose state (on/off) depends on nucleotide binding (ATP/ADP) (Tameling et al. 2006). *NRTP1* encodes a CC-NB-LRR type protein, and semi-dominant mutant *nrt1-D* contains an amino acid substitution in the NB-ARC domain that causes constitutive auto-activation of the NRTP1 protein for a short-root phenotype in rice (Yu et al. 2018). ATP/ADP isopentenyltransferases are likely responsible for most isopentenyladenine- and tZ-type cytokinin synthesis (Miyawaki et al. 2006) and catalyze prenylation of adenosine diphosphate (ADP) or triphosphate (ATP) biosynthesis (Kieber et al. 2014). The hydrolysis of adenosine triphosphate (ATP) is directly coupled to the primary active transport of CTK (Nedvĕd et al. 2021). In peach, ATP/ADP PpIPT genes are key genes for cytokinin biosynthesis in nodal stems (Li et al. 2018).

The NB-ARC domain of *GNP12* protein may play a role in panicle development by influencing hormone activity to control grain number of panicle, length of panicle, and grain length. Future work is required to test this hypothesis and investigate the regulation of these important genes to better develop molecular tools for improved genetic breeding.

Conclusion

In summary, we successfully analyzed NB-ARC family with conserved P-loop domains, identified conservation of NB-ARC orthologous genome segments in monocotyledons. Most NB-ARC genes regulate panicle development in Ce253. And we detected *GNP12* which were controlled panicle length, panicle grain number, and grain length.

Methods

Plant Materials

The rice variety Ce253, widely planted in Guangxi Province, China was selected for this study and obtained from the Rice Research Institute, Guangxi Academy of Agricultural Sciences. Rice was planted under natural field conditions at the Rice Research Institute of Guangxi Academy of Agricultural Sciences, Nanning, China in 2020. The distance between plants within rows was 16 cm, and the distance between plants in separate rows was 20 cm. Field management, including irrigation, fertilizer application, and pest control, were performed according to normal agricultural practices. Fully filled grains were subjected to grain width, length, and weight measurement with a Wanshen SC-G automatic seed test system. All trait measurements were repeated at least three times.

Identification of gene family members

Genome-wide identification of NB-ARC genes from three species of monocotyledonous and dicotyledonous plants was performed. Hidden Markov Model (HMM) (Punta et al. 2000) (version 3.0) analysis was used for the search. HMM profiles of NB-ARC genes (PF00931.2) were obtained from the Pfam database (<http://pfam.xfam.org/>) with an $e\text{-value} \leq 1e-3$. The results of the HMMER sequence alignment were screened to remove protein sequences that were at least 45% longer than the length of the HMM model domain, while retaining the longest protein sequence in the variable shear. Simple Modular Architecture Research Tool (SMART) (Schultz et al. 1998) (version 8.0) (<http://smart.embl-heidelberg.de>) was used for further analysis with all non-redundant protein sequences. Finally, 258 NB-ARC gene models were identified in the rice genome for further analysis. The basic information of the identified NB-ARC proteins was obtained using the tools at the ExPasy website (<http://web.expasy.org/protparam/>). CDS coordinate information is listed in Table S1. Gene family data were analyzed by Gene Denovo Biotechnology Co., Ltd (Guangzhou, China). A BLASTP search of the NCBI nonredundant protein database was used to assign the NB-ARC domains. Multiple alignments of NB-ARC domains with 11 different plant NB-ARC domains of 11 subgroups were performed using MEME. The submotifs were analyzed through <http://weblogo.berkeley.edu/logo.cgi>. The 3D model of the NB-ARC genes was predicted by Jpred and SWISS-MODEL.

Gene structure analysis, chromosomal distribution, and gene duplication

The exon-intron structural information for the collected rice NB-ARC genes was acquired from reference genome annotation files (Os-Nipponbare-Reference-IRGSP-1.0 pseudomolecules) to compare genomic and coding sequences. The cDNAs were aligned with their corresponding genomic DNA sequences. To map all NB-ARC genes, the chromosome distribution and conserved regions were confirmed by analysis of reference genome annotation files. A chromosome distribution diagram was drawn using the SVG package in Perl. Gene duplication data of rice were identified using the replicate gene classifier program of MCScanX software (Wang et al. 2012). All protein-coding sequences were aligned using blastp, and the alignment results were used as input files for MCScanX software to predict gene replication. A gene

was identified as a replication gene according to $e\text{-value} < 1e\text{-}5$ or $e\text{-value} < 1e\text{-}10$. Five replication events were identified: segmental, tandem, proximal, and decentralized (dispersed).

Motifs, phylogenetic, combination diagram, and closely related species analysis

Conserved motifs in the gene family sequences were identified using the MEME program (<http://alternate.meme-suite.org/tools/meme>) with statistical significance (Bailey et al. 2015). The Meme program was run with default settings, maximum motif search value of 15, and an optimum motif width of 10 - 100 amino acid residues. A phylogenetic tree was constructed with the neighbor-joining algorithm in MEGA (version 7.0) with bootstrap test of 1000 times (Kumar et al. 2016). Genome annotations and corresponding protein sequences were downloaded from EnsemblPlants (Brassica_rapa. IVFCAAS v1.36, Brassica_oleracea.v2.1.36, Arabidopsis_thaliana.TAIR10) and GenoScope (Brassica_napus.annotation_v5). The gene structure and motifs were analyzed by systematic evolutionary relationships. Synteny detection was performed by McScanX and drawn with Circos software. A phylogenetic tree was constructed with MEGA (version 7.0) and drawn with iTOL (<https://itol.embl.de/>) to visualize the relationships between the different groups of NB-ARC genes.

Ka/Ks analysis

We compared the Ka/Ks ratios as a proxy of the selective pressures acting on gene pairs for reciprocal best match gene pairs from *A. thaliana*, *Solanum lycopersicum*, *Oryza sativa*, *Triticum aestivum*, and *maize*. Ka/Ks is calculated as the ratio of the number of nonsynonymous substitutions per nonsynonymous site (Ka) in a period to the number of synonymous substitutions per synonymous site (Ks) in the same period. Ka, Ks, and Ka/ Ks values are based on coding sequence alignment and calculated using the KaKs_calculator software package based on the Nei and Gojobori model (Nei et al. 1986). Positively or negatively selected sites were identified based on Ka/Ks ratios with a confidence interval for each ratio given by a P value and an adjusted P value (Adj.Pval) from multiple comparisons (Massingham et al. 2005). Orthologous genes with a Ka/Ks ratio above 1 were under strong positive selection, between 0.5 and 1 were considered to be under weak positive selection, and below 0.1 were considered to be under negative selection (purifying selection).

Vector construction and rice transformation and subcellular localization

To assess subcellular localization, sgRNA-Cas9 plant expression vectors were constructed as described previously (Mao et al. 2013). The targeting sequence (5'- TAGTCGACGACAATGCTGCCAGG-3'), corresponded to +400 to +422 within the third exon encoding the C-terminal end of *GNP12* (starting at amino acid residue 134). To construct the GFP plasmids, the cDNAs of *LOC_Os11g45160*, *LOC_Os12g17490*, *LOC_Os08g42930* were amplified from Nipponbare and the cDNA for *LOC_Os12g36720* was amplified from Ce253. The cDNAs were cloned into the pBWA(V)HS-ccdb-GLogfp vector to generate the forward insertion. Plasmids CaMV35S::GFP, CaMV35S::*LOC_Os101g458510*-GFP and CaMV35S::*GNP12*-GFP were transformed into rice leaf

protoplasts for subcellular localization. GFP was excited with a 488 nm laser and imaging was performed.

DNA extraction, RNA extraction, expression analysis, and RNA-Seq

Panicles of Ce253 of different lengths (3, 5, 10, 15, 20, and 25 cm) at the booting stage were flash-frozen in liquid nitrogen. Total RNA was extracted from these tissues using a EasyPure® Plant RNA Kit (Trans, Catalog no. ER301-01). Total RNA (2.0 µg) was used for cDNA synthesis with a PrimeScript™ RT Reagent Kit with gDNA Eraser (TransScript® II One-Step RT-PCR SuperMix, Tran, Catalog no. AH411-02). The resulting cDNA samples were diluted five-fold and used as templates for qRT-PCR using a TransStart® Green qPCR SuperMix and a CFX96 RealTime system (Bio-RAD, Hercules, CA, USA) following the manufacturer's instructions. The qRT-PCR reactions were performed as 10-µL mixtures containing 5 µL of 2× Green qPCR MasterMix, 1 µL of cDNA, 0.25 µL of each primer (10 µM), and 3.5 µL of ddH₂O. Amplification steps were 95 °C for 30 s, 40 cycles of 95 °C for 5 s, and 60 °C for 30 s, followed by 65 °C for 5 s, 95 °C for 15 s, 60 °C for 30 s, and 95 °C for 15 s. Each experiment was repeated at least three times. The qRT-PCR analysis was performed using the $\Delta\Delta C_t$ method. Details on gene-specific primers used for real-time PCR are provided in Table S1. The ubiquitin gene (*LOC_Os03g13170*) (Chen et al. 2019) was used as a control (* $P < 0.05$; ** $P < 0.01$; Student's t -test). RNA samples used for RNA-seq analysis were prepared from different panicles of Ce253 grown under normal field conditions with three biological replicates. RNA library sequencing was performed on an Illumina Hiseq™ 2500/4000 platform by Majorbio. Sequence analysis was performed using the method provided by Majorbio (<http://www.majorbio.com/>). RNA-Seq data for NB-ARC genes in different tissues were obtained from <http://expression.ic4r.org/>.

Haplotype and evolutionary analysis

The single-nucleotide polymorphisms (SNPs) of 827 accessions used for haplotype analysis were acquired from the 3K rice genomes (3K-RG) dataset. The grain length data for 827 accessions were downloaded from the Rice SNP-Seek Database. One-way ANOVA followed by Duncan's new multiple-range test were performed using SPSS 21.0 software. The genomic sequences of 1400 cultivated and 58 wild accessions were obtained from the 3K rice genomes (3K-RG) dataset and OryzaGenome (<http://viewer.shigen.info/oryzagenome/>), respectively, and were used to construct a minimum spanning tree for RGH1A. Arlequin version 3.5 software was used to calculate a haplotype network and the distance matrix output was used in Hapstar-0.6 to draw a minimum spanning tree. The average nucleotide diversity (π) and Tajima's D for each subpopulation in RGH1A and 40-kb flanking regions were calculated using DnaSP 5.10 software (Librado et al. 2009). The nucleotide diversity curves were generated using 60-bp window and 15-bp step length.

Abbreviations

Chr

chromosome
CTK
Cytokinin
GNP
Grain Number per Panicle
IAA
Indole-3-acetic acid
MEME
Multiple EM for motif elicitation
NB-ARC
nucleotide-binding adaptor shared by APAF-1, R proteins and CED-4
ORF
Open reading frame
qRT-PCR
quantitative Reverse-Transcription Polymerase Chain Reaction
SNPs
Single-nucleotide polymorphisms.

Declarations

Authors' contributions

YHP, YZ and GFD designed the research. LC, RF, CCH and DJQ performed the experiments. YHP, HFG, QJL, MARR, XYZ, HFL and LJG analyzed the data. YHP and LC wrote and revised the manuscript. All authors read and approved the final manuscript.

Funding

This work was supported by the National Natural Science Foundation of China (32060452, 32160447, 32160645, 31760428, U20A2032), the key project of the Guangxi Natural Science Foundation (2018GXNSFDA281053), Natural Science Foundation of Shandong Province (ZR2020MC096), the project of the central government guides the development of local science and technology (Guike ZY20198015, Guike AD17129064, Guike ZY19183020 and Guike ZY18057004), the project of the base and talent (GuiKe 2018JZ24 and 2018-15-Z06-KF10).

Availability of data and materials

All data generated or analyzed in this study are included in this published article and its additional information files.

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

References

1. Andersen EJ, Lindsey LE, Nepal MP (2020) Genome-wide identification of disease resistance genes (R genes) in wheat. bioRxiv <https://doi.org/10.1101/2020.07.18.210286>.
2. Albrecht M, Takken FLW (2006) Update on the domain architectures of NLRs and R proteins. Biochem Biophys Res Commun 339(2):459–462.
3. Ashikari M, Sakakibara H, Lin S, Yamamoto T, Takashi T, Nishimura A, Angeles ER, Qian Q, Kitano H, Matsuoka M (2005) Cytokinin oxidase regulates rice grain production. Science 309(5735):741–745.
4. Bailey TL, Johnson J, Grant CE, Noble WS (2015) The MEME suite. Nucleic Acids Res 43(W1):W39–49.
5. Chandra S, Kazmi AZ, Ahmed Z, Roychowdhury G, Kumari V, Kumar M, Mukhopadhyay K (2017) Genome-wide identification and characterization of NB-ARC resistant genes in wheat (*Triticum aestivum* L.) and their expression during leaf rust infection. Plant Cell Rep 36(7):1097–1112.
6. Chattopadhyaya R, Pal A (2008) Three-dimensional Models of NB-ARC Domains of Disease Resistance Proteins in Tomato, Arabidopsis, and Flax. J Biomol Struct Dyn 25(4):357–371.
7. Chen A, Gu L, Li M, Xu Na, Feng Fajie, Liu Guangliang, Zhang Bao, Gong Daping, Zhang Junyi, Liu Hongyan, Zhang Zhongyi (2018) Identification of *Rehmannia glutinosa* L. NB-ARC family proteins and their typical changes under consecutive monoculture stress. Acta Physiol Plant 40:95.
8. Chen W, Cheng Z, Liu L, Wang M, You X, Wang J, Zhang F, Zhou C, Zhang Z, Zhang H, You S, Wang Y, Luo S, Zhang J, Wang J, Wang J, Zhao Z, Guo X, Lei C, Zhang X, Lin Q, Ren Y, Zhu S, Wan J (2019) Small Grain and *Dwarf 2*, encoding an HD-Zip II family transcription factor, regulates plant development by modulating gibberellin biosynthesis in rice. Plant Sci 288:110208.
9. Deng Y, Chen H, Zhang C, Cai T, Zhang B, Zhou S, Fountain JC, Pan RL, Guo B, Zhuang WJ (2018) Evolution and characterisation of the AhRAF4 NB-ARC gene family induced by *Aspergillus flavus* inoculation and abiotic stresses in peanut. Plant Biol (Stuttg) 20(4):737–750.
10. Dolezal AL, Shu X, O'Brien GR, Nielsen DM, Woloshuk CP, Boston RS, Payne GA (2014) *Aspergillus flavus* infection induces transcriptional and physical changes in developing maize kernels. Front. Microbiol 5:384.
11. Guo T, Lu ZQ, Shan JX, Ye WW, Dong NQ, Lin HX (2020) *ERECTA1* acts upstream of the OsMKKK10-OsMKK4-OsMPK6 cascade to control spikelet number by regulating cytokinin metabolism in rice. Plant Cell 32(9):2763–2779.
12. He H, Ji J, Li H, Tong J, Feng Y, Wang X, Han R, Bie T, Liu C, Zhu S (2020) Genetic diversity and evolutionary analyses reveal the powdery mildew resistance gene *Pm21* undergoing diversifying selection. Front Genet 11:489.

13. Igari K, Endo S, Hibara K, Aida M, Sakakibara H, Kawasaki T, Tasaka M (2008) Constitutive activation of a CC-NB-LRR protein alters morphogenesis through the cytokinin pathway in *Arabidopsis*. *Plant J* 55(1):14–27.
14. Ishimaru K, Hirotsu N, Madoka Y, Murakami N, Hara N, Onodera H, Kashiwagi T, Ujiie K, Shimizu B, Onishi A, Miyagawa H, Katoh E (2013) Loss of function of the IAA-glucose hydrolase gene *TGW6* enhances rice grain weight and increases yield. *Nat Genet* 45(6):707–711.
15. Jacquemin J, Ammiraju JS, Haberer G, Billheimer DD, Yu Y, Liu LC, Rivera LF, Mayer K, Chen M, Wing RA (2014) Fifteen million years of evolution in the *Oryza* genus shows extensive gene family expansion. *Mol Plant* 7(4):642–656.
16. Kieber JJ, Schaller GE (2014) Cytokinins. *Arabidopsis Book* 12:e0168.
17. Kumar S, Stecher G, Tamura K (2016) MEGA7: molecular evolutionary genetics analysis version 7.0 for bigger datasets. *Mol Biol Evol* 33(7):1870–1874.
18. Leipe DD, Koonin EV, Aravind L (2004) STAND, a class of P-loop NTPases including animal and plant regulators of programmed cell death: multiple, complex domain architectures, unusual phyletic patterns, and evolution by horizontal gene transfer. *J Mol Biol* 343(1):1–28.
19. Librado P, Rozas J (2009) DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. *Bioinformatics* 25(11):1451–1452.
20. Li M, Wei Q, Xiao Y, Peng F (2018) The effect of auxin and strigolactone on ATP/ADP isopentenyltransferase expression and the regulation of apical dominance in peach. *Plant Cell Rep* 37(12):1693–1705.
21. Liu E, Liu Y, Wu G, Zeng S, Tran Thi TG, Liang L, Liang Y, Dong Z, She D, Wang H, Zaid IU, Hong D (2016) Identification of a candidate gene for panicle length in rice (*Oryza sativa* L.) via association and linkage analysis. *Front Plant Sci* 7:596.
22. Lukasik E, Takken FL (2009) STANDing strong, resistance proteins instigators of plant defence. *Curr Opin Plant Biol* 12(4):427–436.
23. Mao Y, Zhang H, Xu N, Zhang B, Gou F, Zhu JK (2013) Application of the CRISPR-Cas system for efficient genome engineering in plants. *Mol Plant* 6(6):2008–2011.
24. Massingham T, Goldman N (2005) Detecting amino acid sites under positive selection and purifying selection. *Genetics* 169(3):1753–1762.
25. Michael Weaver L, Swiderski MR, Li Y, Jones JD (2006) The *Arabidopsis thaliana* TIR-NB-LRR R-protein, *RPP1A*; protein localization and constitutive activation of defence by truncated alleles in tobacco and *Arabidopsis*. *Plant J* 47(6):829–840.
26. Miyawaki K, Tarkowski P, Matsumoto-Kitano M, Kato T, Sato S, Tarkowska D, Tabata S, Sandberg G, Kakimoto T (2006) Roles of *Arabidopsis* ATP/ADP isopentenyltransferases and tRNA isopentenyltransferases in cytokinin biosynthesis. *Proc Natl Acad Sci U S A* 103(44):16598–16603.
27. Mondragón-Palomino M, Stam R, John-Arputharaj A, Dresselhaus T (2017) Diversification of defensins and NLRs in *Arabidopsis* species by different evolutionary mechanisms. *BMC Evol Biol* 17(1):255.

28. Nedvěd D, Hošek P, Klíma P, Hoyerová K (2021) Differential subcellular distribution of cytokinins: how does membrane transport fit into the big picture? *Int J Mol Sci* 22(7):3428.
29. Nei M, Gojobori T (1986) Simple methods for estimating the numbers of synonymous and nonsynonymous nucleotide substitutions. *Mol Biol Evol* 3(5):418–426.
30. Ngou BPM, Heal R, Wyler M, Schmid MW, Jones JDG (2022) Concerted expansion and contraction of immune receptor gene repertoires in plant genomes. *bioRxiv*.
<https://doi.org/10.1101/2022.1101.1101.474684>
31. Pal A, Chakrabarti A, Basak J (2007) New motifs within the NB-ARC domain of R proteins: probable mechanisms of integration of geminiviral signatures within the host species of fabaceae family and implications in conferring disease resistance. *J Theor Biol* 246(3):564–573.
32. Pan Y, Chen L, Zhao Y, Guo H, Li J, Rashid MAR, Lu C, Zhou W, Yang X, Liang Y, Wu H, Qing D, Gao L, Dai G, Li D, Deng G (2021) Natural variation in *OsMKK3* contributes to grain size and chalkiness in rice. *Front Plant Sci* 12:784037.
33. Punta M, Coggill PC, Eberhardt RY, Mistry J, Tate J, Boursnell C, Pang N, Forslund K, Ceric G, Clements J, Heger A, Holm L, Sonnhammer EL, Eddy SR, Bateman A, Finn RD (2012) The Pfam protein families database. *Nucleic Acids Res* 40:D290–D301.
34. Schultz J, Milpetz F, Bork P, Ponting CP (1998) *SMART*, a simple modular architecture research tool: identification of signaling domains. *Proc Natl Acad Sci USA* 95(11):5857–5864.
35. Shi CL, Dong NQ, Guo T, Ye WW, Shan JX, Lin HX (2020) A quantitative trait locus *GW6* controls rice grain size and yield through the gibberellin pathway. *Plant J* 103(3):1174–1118.
36. Slootweg E, Tameling W, Roosien J, Lukasik E, Joosten M, Takken F, Bakker J, Goverse A (2009) An outlook on the localisation and structure-function relationships of R proteins in solanum. *Potato Res* 52(3):229–235.
37. Steele JFC, Hughes RK, Banfield MJ (2019) Structural and biochemical studies of an NB-ARC domain from a plant NLR immune receptor. *PLoS ONE* 14(8):e0221226.
38. Tameling WI, Elzinga SD, Darmin PS, Vossen JH, Takken FL, Haring MA, Cornelissen BJ (2002) The tomato R gene products I-2 and MI-1 are functional ATP binding proteins with ATPase activity. *Plant Cell* 14(11):2929–2939.
39. Tameling WI, Vossen JH, Albrecht M, Lengauer T, Berden JA, Haring MA, Cornelissen BJ, Takken FL (2006) Mutations in the NB-ARC domain of I-2 that impair ATP hydrolysis cause autoactivation. *Plant Physiol* 140(4):1233–1245.
40. van der Biezen EA, Jones JD (1998) The NB-ARC domain: a novel signalling motif shared by plant resistance gene products and regulators of cell death in animals. *Curr Biol* 8(7):R226–227.
41. Wang S, Li S, Liu Q, Wu K, Zhang J, Wang S, Wang Y, Chen X, Zhang Y, Gao C, Wang F, Huang H, Fu X (2015) The *OsSPL 16-GW7* regulatory module determines grain shape and simultaneously improves rice yield and grain quality. *Nat Genet* 47(8):949–954.
42. Wang W, Mauleon R, Hu Z, Chebotarov D, Tai S, Wu Z, Li M, Zheng T, Fuentes RR, Zhang F, Mansueto L, Copetti D, Sanciango M, Palis KC, Xu J, Sun C, Fu B, Zhang H, Gao Y, Zhao X, Shen F, Cui X, Yu H,

- Li Z, Chen M, Detras J, Zhou Y, Zhang X, Zhao Y, Kudrna D, Wang C, Li R, Jia B, Lu J, He X, Dong Z, Xu J, Li Y, Wang M, Shi J, Li J, Zhang D, Lee S, Hu W, Poliakov A, Dubchak I, Ulat VJ, Borja FN, Mendoza JR, Ali J, Li J, Gao Q, Niu Y, Yue Z, Naredo MEB, Talag J, Wang X, Li J, Fang X, Yin Y, Glaszmann JC, Zhang J, Li J, Hamilton RS, Wing RA, Ruan J, Zhang G, Wei C, Alexandrov N, McNally KL, Li Z, Leung H (2018) Genomic variation in 3,010 diverse accessions of Asian cultivated rice. *Nature* 557(7703):43–49.
43. Wang Y, Tang H, Debarry JD, Tan X, Li J, Wang X, Lee TH, Jin H, Marler B, Guo H, Kissinger JC, Paterson AH (2012) MCScanX: a toolkit for detection and evolutionary analysis of gene synteny and collinearity. *Nucleic Acids Res* 40(7):e49.
44. Wang Y, Tang J, Li H, Teng Z, Wang W, Xu F, Sun K, Chu J, Qian, Y, Loake GJ, Chu C (2020) A NB-ARC-CRRSP signaling module triggers HR-like cell death and associated disease resistance in rice by suppressing antioxidant defense systems. <http://dx.doi.org/10.2139/ssrn.3565037>
45. Wen Z, Yao L, Singer SD, Muhammad H, Li Z, Wang X (2017) Constitutive heterologous overexpression of a TIR-NB-ARC-LRR gene encoding a putative disease resistance protein from wild Chinese *Vitis pseudoreticulata* in *Arabidopsis* and tobacco enhances resistance to phytopathogenic fungi and bacteria. *Plant Physiol Biochem* 112:346–361.
46. Wu Y, Wang Y, Mi XF, Shan JX, Li XM, Xu JL, Lin HX (2016) The QTL *GNP1* encodes GA20ox1, which increases grain number and yield by increasing cytokinin activity in rice panicle meristems. *PLoS Genet* 12(10):e1006386.
47. Yildirim-Ersoy F, Ridout CJ, Akkaya MS (2011) Detection of physically interacting proteins with the CC and NB-ARC domains of a putative yellow rust resistance protein, Yr10, in wheat. *J Plant Dis Prot* 118(3/4):119–126.
48. Yoshida A, Ohmori Y, Kitano H, Taguchi-Shiobara F, Hirano HY (2012) *ABERRANT SPIKELET AND PANICLE1*, encoding a TOPLESS-related transcriptional co-repressor, is involved in the regulation of meristem fate in rice. *Plant J* 70(2):327–339.
49. Yu Z, Dong L, Jiang Z, Yi K, Zhang J, Zhang Z, Zhu Z, Wu Y, Xu M, Ni J (2018) A semi-dominant mutation in a CC-NB-LRRtype protein leads to a short-root phenotype in rice. *Rice* 11(1):54.
50. Zhao L, Tan L, Zhu Z, Xiao L, Xie D, Sun C (2015) *PAY1* improves plant architecture and enhances grain yield in rice. *Plant J* 83(3):528–536.

Tables

Table 1,2 is available in the Supplemental Files section.

Figures

Figure 1

Phylogenetic tree (a), exon/intron structure (b), and motif composition (c) of NB-ARC genes in rice. Neighbor-Joining (NJ)-Phylogenetic trees shown in a and b were prepared using the same methods. The widths of gray bars at the bottom indicate relative lengths of genes and proteins. Yellow boxes and blue lines in b represent exons and introns, respectively. Different boxes in c represent different motifs. Different background colors represent different groups and subgroups.

Figure 2

Chromosomal distribution of NB-ARC gene duplication events. **a** Distribution of NB-ARC genes with indicated duplication types. **b** Numbers of duplicated genes for different duplication types. **c** Collinearity of replicative genes in the protein family.

Figure 3

Phylogenetic, Ka/Ks, and synteny analysis of NB-ARC genes from rice, *Arabidopsis*, and *Triticum aestivum*. **a** Group I, Group II, Group III, IIa, IIb, IIc, IId, IIe, IIf, IIg, IIh, and IIf were identified in the rice genome; 1052 and 50 NB-ARC genes were identified in *Triticum aestivum* and *Arabidopsis thaliana*. Genes from rice, *Arabidopsis*, *Triticum aestivum* are indicated with the prefixes Os, At, and Tran, respectively. **b** Synteny analysis of *Oryza sativa*, *Arabidopsis*, and *Triticum aestivum*. Gray lines in the background indicate collinear blocks of plant genomes. Different color bars represent the chromosomes of different species. The chromosome number is labeled at the top or bottom of each chromosome. **c** Ka/Ks analysis of NB-ARC genes from rice, *Arabidopsis* and *Triticum aestivum*.

Figure 5

The effect of *GNP12* on cell proliferation length of panicle, grain number of panicle and grain size. **a** Length of panicle of Ce253 and Cr-lines. Scale bars, 5 cm. **b** Length of grains from Ce253 and Cr-line. **c** Width of grains from Ce253 and Cr-line. Scale bars, 1 cm. **d** Scanning electron microscopic images of chalkiness in transgenic plants and Ce253. Bar, 50 μ m. **e-l** Panicle length (cm) (**e**), grain number of panicle (**f**), primary branch number (**g**), setting percentage (%) (**h**), grain of length-width ratio (**i**), grain

length (mm) (j), grain width (mm) (k), CTK content in Ce253 and Cr-line (l). Data are given as mean $\pm$ s.e.m. (n = 12). Student's *t* tests were used to generate *P* values. Haplotype and evolutionary analysis of *GNP12* in rice. m Major haplotypes of *GNP12* using 32 single-nucleotide polymorphisms (SNPs) in the coding sequence region in 827 accessions. n Phenotypic evaluation of haplotype grain length in different subgroups. o Haplotype network analysis of *GNP12*. p Nucleotide diversity of *GNP12*.

Figure 6

Predicted structures and subcellular localization of putative NB-ARC protein domains based on 11 proteins from each 11 subgroups. **a** Conservation of key motifs, residues, and secondary structure between putative NB-ARC domains. Dashed lines indicate positions within secondary structure elements. The histograms above the motifs represent the degree of conservation (% of identical to conserved residuals) for putative NB-ARC domains (blue columns) among 11 proteins from 11 subgroups. **b** Sequence logos representing the conservation of key motifs and neighboring sequences. The size of the letters corresponds to information content. **c** 3D model of 11 proteins from 11 subgroups in NB-ARC family. **d** Subcellular localization of three proteins from groups I, II, and III.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [Table1.xlsx](#)
- [Table2.xlsx](#)
- [TableS1.xlsx](#)
- [TableS2.xlsx](#)
- [TableS3.xlsx](#)
- [TableS4.xlsx](#)
- [TableS5.xlsx](#)
- [TableS6.xlsx](#)