Plant Molecular Biology Reporter
https://doi.org/10.1007/s11105-018-1086-y

ORIGINAL PAPER

Genome-Wide Analysis in Wild and Cultivated *Oryza* Species Reveals Abundance of NBS Genes in Progenitors of Cultivated Rice

Hukam C. Rawal[1] · S. V. Amitha Mithra[1] · Kirti Arora[1] · Vishesh Kumar[1] · Neha Goel[1] · Dwijesh Chandra Mishra[2] · K. K. Chaturvedi[2] · Anil Rai[2] · S. Vimala Devi[3] · T. R. Sharma[1,4] · Amolkumar U. Solanke[1]

Abstract

NBS-encoding genes play a critical role in the plant defense system. Wild relatives of crop plants are rich reservoirs of plant defense genes. Here, we performed a stringent genome-wide identification of NBS-encoding genes in three cultivated and eight wild *Oryza* species, representing three different genomes (AA, BB, and FF) from four continents. A total of 2688 NBS-encoding genes were identified from 11 *Oryza* genomes. All the three progenitor species of cultivated rice, namely *O. barthii*, *O. rufipogon*, and *O. nivara*, were the richest reservoir of NBS-encoding genes (214, 313, and 307 respectively). Interestingly, the two Asian cultivated species showed a contrasting pattern in the number of NBS-encoding genes. While *indica* subspecies maintained nearly equal number of NBS genes as its progenitor (309 and 313), the *japonica* subspecies had retained only two third in the course of evolution (213 and 307). Other major sources for NBS-encoding genes could be (i) *O. longistaminata* since it had the highest proportion of NBS-encoding genes and (ii) *O. glumaepatula* as it clustered distinctly away from the rest of the AA genome species. The present study thus revealed that NBS-encoding genes can be exploited from the primary gene pool for disease resistance breeding in rice.

Keywords NBS genes · NB-ARC domain · NBS-LRR · Genome-wide analysis · Disease resistance · Evolution

Introduction

Plant disease development is a continuous war between plants and pathogens wherein win by plants results in disease resistance while defeat leads to plant susceptibility. Plant immune system, evolved and stabilized to protect itself from all types of pathogens, can be divided into two categories. The pathogen-associated molecular pattern (PAMP)-triggered immunity (PTI) is the first line of defense activated by flagellin- or chitin-like pathogen patterns forming the basis of broad-spectrum resistance. However, to suppress PTI, many pathogens produce effector molecules, which help them to infect plants. To avoid this effector-mediated entry of pathogens, plants have evolved a second line of immune system, effector-triggered immunity (ETI) (Liu et al. 2014). When effector proteins enter the cell, they are encountered by resistance (R) proteins which in turn activate the defense system. Most of these R proteins are NBS-LRR (nucleotide binding site-leucine-rich repeats) domains containing proteins, sometimes with TIR (toll and interleukin-1 receptor) or CC (coiled-coil) motifs at N-terminal (Yue et al. 2012). In these proteins, the NBS domain is responsible for ATP or GTP binding and hydrolysis, whereas LRR domains are involved in protein-protein interaction and binding to pathogen-derived molecules (Van der Biezen and Jones 1998; Lukasik and Takken 2009). NBS proteins without LRRs are also found to have role in defense response (Nandety et al. 2013).

R proteins are abundant in plant species where each gene is specific to a particular pathogen effector (Li et al. 2010; Singh et al. 2015; Seo et al. 2016). In dicots, 159 NBS genes

Hukam C. Rawal and S. V. Amitha Mithra contributed equally to this work.

Electronic supplementary material The online version of this article (https://doi.org/10.1007/s11105-018-1086-y) contains supplementary material, which is available to authorized users.

✉ Amolkumar U. Solanke
amol.solanke@icar.gov.in

[1] ICAR-National Research Centre on Plant Biotechnology, Pusa Campus, New Delhi 110012, India

[2] ICAR-Indian Agricultural Statistics Research Institute, New Delhi, India

[3] ICAR-National Bureau of Plant Genetic Resources, New Delhi, India

[4] National Agri-Food Biotechnology Institute, Mohali, India

Published online: 12 May 2018

in *Arabidopsis*, 333 in *Medicago*, 459 in grapevine, 54 in papaya, 228 in cassava, 294 in tomato, 443 in pepper, and 755 in potato have been reported (Guo et al. 2011; Andolfo et al. 2014; Ameline-Torregrosa et al. 2008; Porter et al. 2009; Lozano et al. 2015). Similarly, in monocots, 129 NBS genes in maize, 245 in *sorghum*, 239 in *brachypodium*, and 330 in poplar are known (Li et al. 2010; Yang et al. 2008). Variation in the number of R proteins within same species has also been observed due to different parameters, stringencies, and genotypes used for identification. In all these species, NBS-encoding genes are found distributed into large and diverse families which are present in clusters in the genome (Joshi and Nayak 2013). The structural diversity of NBS gene clusters and development of specific resistance are due to gene duplication, unequal crossing over, ectopic recombination, and diversifying selection which resulted in rapid NBS gene evolution (Singh et al. 2015). The number of NBS-encoding genes varies even in closely related species. Its number is proportionate neither to genome size nor to the number of all predicted genes in the genome of the species. This is due to rapid expansion or contraction and the speedy structural evolution in these loci caused by natural selection pressure of quickly changing species-specific pathogen spectrum (Li et al. 2010).

Wild species of agronomically important crops always remain a potential source of unexploited genes and allelic variations (Sanchez et al. 2014). Since half of the world is dependent on rice for its carbohydrate need, the use of gene prospecting and allele mining from wild relatives of rice is a valuable approach for search of superior genes and alleles. The genus *Oryza* has two domesticated and 22 wild species representing the six diploid and four tetraploid genomes (Vaughan 1994). Many yield-related and biotic and abiotic stress tolerance genes have been introgressed into cultivated rice from wild relatives (Nguyen et al. 2013; Ram et al. 2007; Rahman et al. 2009). Many resistant genes for bacterial blight resistance (*Xa21*, *Xa23*, *Xa27*, *Xa29(t)*, and *Xa38*) and blast resistance (*Pi9*, *Pi40*, *Pi54of*, *Pi54rh*) are from wild species of rice like *O. nivara*, *O. rufipogon*, *O. longistaminata*, *O. officinalis*, *O. minuta*, *O. rhizomatis*, and *O. australiensis* (Sanchez et al. 2014; Das et al. 2012; Devanna et al. 2014). Though comparison of wild rice species based on chloroplast genomes, domestication, and genome-wide comparison of AA genome species have been carried out, in-depth identification of NBS genes from wild rice species remains unexploited (Molina et al. 2011; Zhang et al. 2014; Wambugu et al. 2015). With the availability of genome sequences of nine wild rice species, we carried out the present study to identify all NBS genes from three cultivated and eight wild rice species representing three diploid genomes, and analyzed their chromosomal positions, gene clusters on genomes, conservation among species, species specificity, and evolutionary pattern to exploit best NBS genes for disease resistance in rice.

Materials and Methods

Resources

Protein, genomic, and chromosome sequences of the 11 *Oryza* species, namely, *O. sativa* ssp. *japonica* (cv. Nipponbare), *O. sativa* ssp. *indica* (93-11), *O. rufipogon* (W1943), *O. punctata* (IRGC105690), *O. nivara* (IRGC100897), *O. meridionalis* (OR44), *O. longistaminata* (IRGC110404), *O. glumaepatula* (IRGC105668), *O. glaberrima* (IRGC96717), *O. brachyantha* (IRGC101232), and *O. barthii* (IRGC105608), were downloaded from Gramene (ftp://ftp.gramene.org/ accessed on 09/20/2016) (Table 1). Of these, *O. sativa* ssp. *japonica*, *O. sativa* ssp. *indica*, and *O. glaberrima* are the cultivated species and the rest eight are wild relatives of rice (Sanchez et al. 2014). For validation of the identified genes using PCR, accessions belonging to five different species were procured from ICAR-National Bureau of Plant Genetic Resources, New Delhi.

Identification of NBS-Encoding Genes

A methodology was developed to predict the NBS-encoding genes, based on the previous studies in *Arabidopsis*, cassava, and Solanaceae family with few modifications (Seo et al. 2016; Guo et al. 2011; Lozano et al. 2015; Meyers et al. 2003). Downloaded protein sequences of the 11 *Oryza* species were scanned by HMM search module of HMMER v3.0 package (http://hmmer.org/) against Pfam NBS family (PF00931) (http://pfam.xfam.org/). Significant hits with *E* value threshold of $\leq e^{-60}$ were extracted, aligned, and used to build an *Oryza*-specific NBS HMM profile with the "hmmbuild" tool of HMMER package. All protein sequences were scanned against this newly built *Oryza*-specific NBS HMM profile and Pfam NBS family (PF00931) by HMMsearch (*E* value threshold of $\leq e^{-60}$) and against Pfam30.0 with InterProScan 5.20 (*E* value threshold of $< e^{-5}$) to identify the "Candidate NBS genes." (Jones et al. 2014) The genes present in both the first list and Candidate NBS genes list were considered the "Predicted NBS genes" while those present in any one of these two lists were named as "Putative NBS genes." The redundant sequences were removed and the longest isoforms were kept for further analysis.

Chromosomal Location of NBS Genes

Chromosome sequences were downloaded from Gramene (ftp://ftp.gramene.org/) for each of the *Oryza* species under study except *O. longistaminata* and *O. meridionalis*, as their chromosome-wise data were not available. For these two species, the predicted NBS genes were BLAST searched against the chromosome sequences of *O. sativa* ssp. *japonica*

Table 1 Number of NBS genes predicted across the 11 *Oryza* species

Oryza sp.	Cultivar	Genome type	Genome size (Mb)	Total genes	Predicted NBS genes (best isoforms)
O. sativa ssp. *japonica*	Nipponbare	AA	374.42	42,132	213
O. sativa ssp. *indica*	93–11	AA	411.71	40,745	309
O. rufipogon	W1943	AA	338.04	47,441	313
O. punctata	IRGC105690	BB	393.82	41,060	207
O. nivara	IRGC100897	AA	337.95	48,360	307
O. meridionalis	OR44	AA	335.66	43,455	97
O. longistaminata	IRGC110404	AA	326.44	31,686	269
O. glumaepatula	IRGC105668	AA	372.86	46,893	293
O. glaberrima	IRGC96717	AA	316.42	33,164	182
O. brachyantha	IRGC101232	FF	260.84	32,037	184
O. barthii	IRGC105608	AA	308.27	41,595	314
Total				448,468	2688

to predict their possible chromosomal locations. For the rest of the nine species, the predicted NBS genes were BLAST searched against the chromosome sequences of their respective genomes using BLAST 2.3.0+ to obtain their physical location on the chromosome (Altschul et al. 1990; Camacho et al. 2009).

Predicting Domain Patterns

Predicted NBS genes were subjected to the CD-search tool v3.15 at NCBI against Pfam database (http://pfam.xfam.org) with default parameters to predict all the different domains present and their pattern of distribution in the gene sequence (Marchler-Bauer et al. 2011). These results were also used to confirm the presence of NB-ARC domain in all of the predicted NBS genes.

Sequence Alignment and Phylogenetic Analysis

Predicted NBS genes in the 11 *Oryza* species were aligned using ClustalX2.1 and used to construct a phylogenetic tree by the neighbor-joining method with 1000 bootstrap replicates (Larkin et al. 2007).

Orthologous, Paralogous, and Species-Specific NBS Genes

Orthologous and paralogous pairs were detected among the predicted NBS genes by an "all against all BLAST search"-based approach with higher identity percentage similar to previous studies (Hulsen et al. 2006; Heidel et al. 2011; Rawal et al. 2013). If any two significant hits belonging to two different species had at least 40% identity over 70% of protein length with a bit score ≥ 100 and E value $\leq 1e^{-20}$ and were bidirectional hits with each other, these were considered an orthologous pair. Any bidirectional hits from the same species with a bit score ≥ 100 and E value $\leq 1e^{-20}$ and at least 75% identity were considered paralogous pairs. These paralogous pairs were then analyzed at three stringency levels for ancient to recent duplication events: low stringency (for more ancient duplication, ~40 MYA) where at least 75% length of the smaller gene is identical to $\geq 50\%$ length of the larger gene of a paralogous pair; medium stringency (ancient duplication, ~6 MYA) with 85 and 70% identity level; and as high stringency (recent duplication, ~2 MYA) with threshold of 95 and 80% (Van Zee et al. 2016). Genes with no orthologous pairs were considered species-specific NBS genes. Species-specific NBS genes were compared and BLAST searched against their respective chromosome sequences to obtain their chromosomal location. For visualization of these chromosomal locations, a chromosome map was generated by MapChart 2.3 (Voorrips 2002).

Results and Discussion

Sequence Resources

The 11 *Oryza* species selected for identification of NBS-encoding genes and their comparisons included *O. sativa* ssp. *japonica*, *O. sativa* ssp. *indica*, *O. rufipogon*, *O. punctata*, *O. nivara*, *O. meridionalis*, *O. longistaminata*, *O. glumaepatula*, *O. glaberrima*, *O. brachyantha*, and *O. barthii*, representing three different genomes (AA, BB, and FF) and four continents (Asia, Africa, Australia, and South America). From the downloaded protein, genomic, and chromosome sequences, we could find no direct

correlation between the number of genes (peptide sequences) and the reported genome size (Table 1). These results are similar with earlier finding, where genome size was not correlated with the number of NBS genes in gramineous species (Li et al. 2010). Here, *O. nivara* had the highest number of genes (48,360) followed by *O. rufipogon* (47,441) and *O. glumaepatula* (46,893), while *O. sativa* ssp. *indica* had the largest genome (411.71 Mb) followed by *O. punctata* (393.82 Mb) and *O. sativa* ssp. *japonica* (374.42 Mb). Thus, wild ancestral species had more number of genes as compared to the cultigens. Such a pattern could not be observed with respect to their genome size.

Identification of NBS-Encoding Genes and Their Confirmation Using Cloned R Genes

HMMsearch followed by InterProScan-based approach (Fig. S1) was adopted to predict the NBS-encoding genes employing more stringent criteria as compared to previous studies to overcome the issues related with irregular quality of the considered genome sequences (Seo et al. 2016; Guo et al. 2011; Lozano et al. 2015; Meyers et al. 2003). A total of 3334 candidate NBS genes were identified across the 11 *Oryza* species against *Oryza*-specific HMM for NBS domain at *E* value threshold of $\leq e^{-60}$ (Table S1). Of these, 2688 qualified as the predicted NBS genes (Table 1, Table S1). The number of predicted NBS-encoding genes had no direct correlation with the total number of genes in the respective *Oryza* species. *O. barthii* had the highest number of NBS-encoding genes (314) followed by *O. rufipogon* (313), *O. sativa* ssp. *indica* (309), and *O. nivara* (307). It was interesting to note that *indica*, the most important cultigen and subspecies, almost had the same number of NBS genes as that of its progenitor species, *O. rufipogon* and *O. nivara*, while the other subspecies, *japonica*, had lost nearly one third of the NBS genes. An earlier BLAST-based identification of R genes in three *Oryza* species reported *japonica* to have the maximum number of R genes (786) followed by *indica* (672) and *brachyantha* (251) with majority of them (45 to 51%) belonging to LRR category (Singh et al. 2015). The selection criteria employed in their study was keyword/phrase-based one as against the more comprehensive bioinformatic approach used in our study focusing only on NBS genes. Though we expected wide variation in the fraction of NBS genes as against the total number of genes across the rice species (Table 1 and Table S1), owing to the stringent criteria used by us for predicting them, the variation was minimal, ranging from 0.22% in *O. meridionalis* to 0.85% in *O. longistaminata* with intermediates like *O. punctata* (0.50%). This is quite comparable with the findings in other organisms like *Arabidopsis* (0.81%) (Meyers et al. 2003), *Manihot esculenta* (1.06%) (Lozano et al. 2015), *Medicago truncatula* (0.65%) (Ameline-Torregrosa et al. 2008), poplar (0.91%) (Tuskan et al. 2006), grape (1.75%) (Jaillon et al. 2001), sorghum (0.18%) (Cheng et al. 2009), and papaya (0.22%) (Porter et al. 2009). Moreover, the results suggested contraction in the number of NBS-encoding genes from the AA genome species (314 to 269 genes) with exceptions of two cultivated species, *O. glaberrima* (182) and the subspecies *japonica* (213) and a wild relative *O. meridionalis* (97) to BB genome, *O. punctata* (207 genes) and to FF genome, *O. brachyantha* (184 genes).

The quality of the nucleotide sequences and their extent of coverage of the whole genome in each species could also have contributed to the differences in the gene prediction among them. Since we have used protein sequences for gene prediction in the study, such discrepancies might have still crept in. However, we expect stringent criteria used by us for the identification of NBS-encoding genes would have taken care of the poor quality though not of coverage. To ascertain this and to add biological significance to the study, we randomly selected 19 of the available cloned, characterized, and mapped rice R genes and downloaded their respective sequences from NCBI (Sharma et al. 2012; Tanweer et al. 2015). Among the predicted 2688 NBS-encoding genes identified in the study, we found hits for all of them in one or more species except *Pi21*, which was the shortest one (266 amino acids) among the selected R genes (Table 2). Nine genes were found to be present in all the 11 species while one of them was absent in *O. glaberrima* (*Pid2*) and eight of them were missing in *O. meridionalis*.

Chromosome-Wise Distribution of NBS Genes

Of the 2688 predicted NBS genes, the maximum number of genes was located on chromosome 11 (591) followed by chromosomes 12 (312), 8 (280), and 1 (245) (Fig. 1, Table S2). Chromosomes 11 and 12 harbored the highest number of NBS genes in *O. barthii* while chromosome 8 and chromosome 1 had the highest number in *O. rufipogon* and *indica* subspecies, respectively. Across all the species, NBS genes were unevenly distributed among all 12 chromosomes but with relatively high density, ranging from 17.80% (in *indica*) to 24.41% (in *japonica*), on chromosome 11. Relatively fewer NBS genes were present on chromosomes 7, 3, 5, 9, and 10 across all the species studied.

Domains and Domain Pattern Prediction in NBS Genes

Domains and their patterns were detected for each of the predicted NBS-encoding genes by the CD-search tool against Pfam database (Fig. 2). Different combinations of

Table 2 Confirmation of the NBS-encoding genes predicted in the study using 19 cloned and characterized rice R genes through BLAST search

S. no.	Gene	Identified from Indica (I) or Japonica (J)	Chromosome	Amino acid length	Domains	BLAST search result with 2688 predicted NBS-encoding genes
1.	*Pi21*	J	4	266	No domain predicted	No hits found
2.	*Pi-kh*	I	11	330	LRR_3	Hits in all 11 species
3.	*Pid-2*	I	6	845	Pkinase_Tyr-Pkinase	Hits with all except *Oryza glaberrima*
4.	*Pid3*	I	6	924	RX-CC_like-LRR_3-NB-ARC-LRR-LRR_8	Hits in all 11 species
5.	*Pita*	J	12	928	AAA_16-P-loop_NTPase-NB-ARC	Hits in all 11 species
6.	*Pit*	J	1	989	RX-CC_like-NB-ARC-LRR-LRR_8-LRR_3	Hits in all 11 species
7.	*Pi-2*	I	6	1032	RX-CC_like-NB-ARC-AAA_16-P-loop_NTPase	Hits with all except *Oryza meridionalis*
8.	*Pi9*	I	6	1032	RX-CC_like-NB-ARC-	Hits with all except *Oryza meridionalis*
9.	*Piz-t*	J	6	1033	RX-CC_like-NB-ARC-AAA_16-P-loop_NTPase	Hits with all except *Oryza meridionalis*
10.	*Pi36*	I	8	1056	RX-CC_like-LRR_3-NB-ARC	Hits in all 11 species
11.	*Pi5*	J	9	1063	NB-ARC-LRR_3-LRR-LRR_RI-LRR_8	Hits in all 11 species
12.	*Pia*	J	11	1116	RX-CC_like-NB-ARC-AAA_22-P-loop_NTPase-LRR-LRR_8	Hits in all 11 species
13.	*Pik-p*	J	11	1142	RX-CC_like-NB-ARC-LRR-LRR_8	Hits with all except *Oryza meridionalis*
14.	*Pi-k*	J	11	1143	RX-CC_like-LRR_3-NB-ARC-LRR-LRR_8	Hits with all except *Oryza meridionalis*
15.	*Pikm*	J	11	1143	RX-CC_like-LRR_3-NB-ARC-LRR-LRR_8	Hits with all except *Oryza meridionalis*
16.	*Pib*	J	2	1251	RX-CC_like-NB-ARC-AAA_16-P-loop_NTPase-LRR_3	Hits with all except *Oryza meridionalis*
17.	*Pi37*	J	1	1290	NB-ARC-LRR-LRR_8	Hits in all 11 species
18.	*Pis-h*	I	1	1290	NB-ARC-LRR-LRR_8	Hits in all 11 species
19.	*Pb1*	J	11	1296	RX-CC_like-NB-ARC	Hits with all except *Oryza meridionalis*

domain patterns were observed and about 60% of these genes (1616) included at least one or other of the three domains viz., "AAA," "LRR," and "P-Loop_NTPase" besides NB-ARC domain. In some predicted NBS genes, additional domains such as "jacalin," "thioredoxin," "reductase," "WRKY," "WD40," "DUF," and "PPR" were also present. We classified the 2688 NBS-encoding genes into 10 different classes as: NBS, NBS-LRR, LRR-NBS, NBS-X, X-NBS, X-NBS-LRR, X-NBS-LRR-X, NBS-LRR-X, NBS-X-LRR, and X-NBS-X-LRR based on the presence or

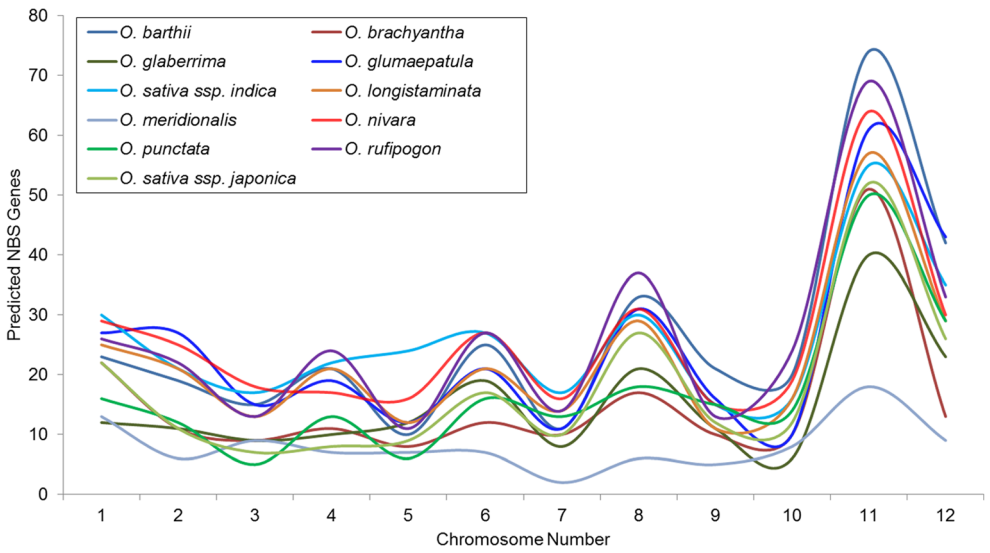

Fig. 1 Chromosome-wise distribution of predicted NBS genes. Sequences of predicted NBS genes in *O. longistaminata* and *O. meridionalis* were aligned against *O. sativa* ssp. *japonica* chromosome sequences to find out their distribution and positions as chromosome-wise data is not available for these two species

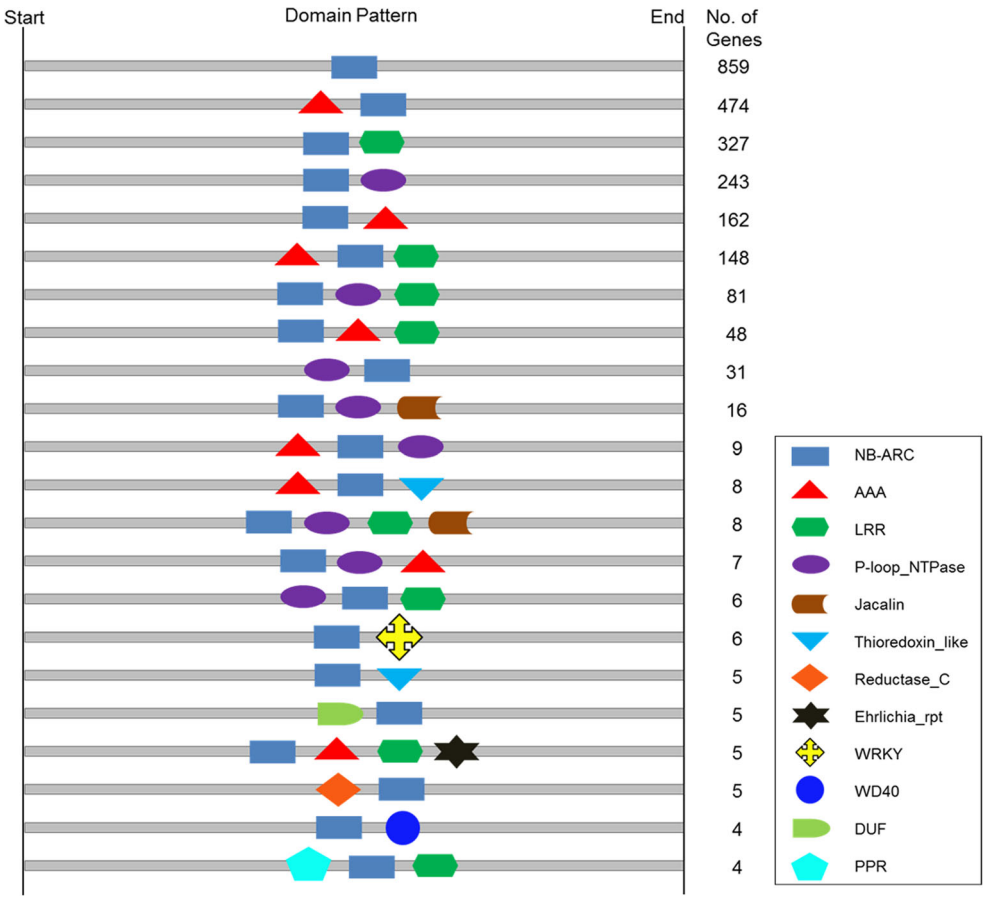

Fig. 2 Showing top 25 domain patterns (in respect to genes following that pattern), irrespective of their copies and position in the genes but in order of their first occurrence in the genes as determined by the CD-search tool at NCBI. The numbers of genes following that pattern are shown on the right side. The patterns with at least four genes following that pattern are shown here for a total of 2535 genes while 116 patterns for the remaining 153 genes are available in supplementary file 3

absence as well as its distribution of LRR domain, where X represented any domain other than the NBS and LRR (Table S3). More than one fourth of the NBS genes (694 genes; 25.82%) had both NB-ARC and LRR domains (Excel Data Sheet 1). The distribution pattern observed for the predicted NBS-encoding and NBS-LRR domain containing proteins was similar in all 11 *Oryza* species (Table S4, Fig. S2). Similar to NBS-encoding genes, NBS-LRR genes also showed reduction in their number from the AA genome with 60 to 85 genes (with exceptions of 43 genes in *O. glaberrima* and 28 genes in *O. meridionalis*) to 57 genes in BB genome and 47 genes in FF genome.

Sequence Alignment and Phylogenetic Analysis

For this large and diverse set of all the predicted NBS-encoding genes in the 11 *Oryza* species, we performed Multiple Sequence Alignment (MSA) and tree generation with the software like MAFFT and Muscle at CIPRES Science Gateway (https://www.phylo.org/) but the best representation was observed with the neighbor-joining (NJ) tree constructed using ClustalX2.1 with 1000 bootstrap replicates. There were two main clades in the NJ tree, one smaller clade consisting of *O. sativa* ssp. *japonica* and *O. rufipogon* and the other larger clade with the remaining eight species (Fig. 3). This larger clade had two sub-clades, one consisting of the Australian (*O. meridionalis*) and African (*O. barthii*) wild species while four South Asian and an African AA genome *Oryza* species were grouped together consisting of the cultivated *O. glaberrima* and *O. sativa* ssp. *indica* and the ancestral species, *O. longistaminata* and *O. nivara.* The species representing BB (*O. punctata*) and FF (*O. brachyantha)* genomes emerged as two separate outer clades in the AA genome clade. In our study, the South American AA genome rice species, *O. glumaepatula*, formed a separate clade and thus acted as an out-group. Our phylogeny results, based on NBS genes alone, were different from the one obtained by Zhang et al. where the genome-wide comparison of five AA genome species had the FF genome wild relative *O. brachyantha* as an outlier (Zhang et al. 2014).

Had the basis of clustering been strictly domestication or evolution, through chloroplast genome or genome-wide markers, we would have expected a clustering that would reflect the progenitor and descendent relationships, geographical proximities, and genomic similarities as found in earlier studies (Sanchez et al. 2014; Molina et al. 2011; Zhang et al. 2014; Wambugu et al. 2015; He et al. 2011). We attribute the drastic differences in clustering between the earlier studies and ours to two major reasons: (1) the inclusion of *O. longistaminata* (AA genome) in the present study which had the highest proportion of NBS genes

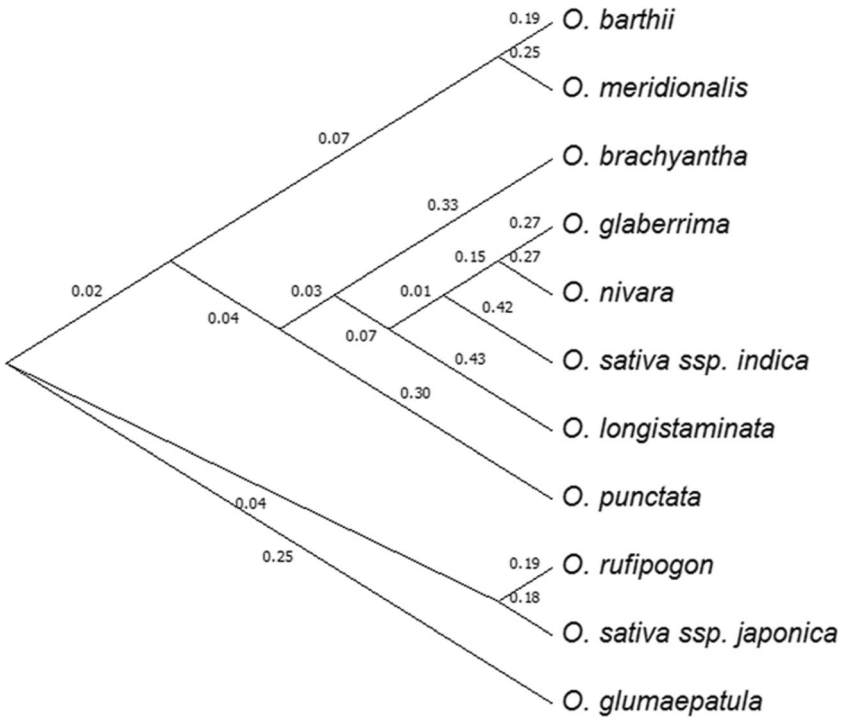

Fig. 3 A guide tree generated by ClustalX 2.1 based on the alignment of all the predicted NBS genes showing the evolutionary relation between 11 *Oryza* species

among all the 11 species compared to the study on AA genome harboring *Oryza* species (Table 1)(Zhang et al. 2014) and (2) focus on the NBS gene family alone which would have evolved in response to the pathogen's presence and load as against the entire gamut of nuclear or chloroplast genomes.(Sanchez et al. 2014; Zhang et al. 2014; Wambugu et al. 2015)

Our analysis observed similarity with the evolutionary history of *Oryza* species in distinguishing the South and Central American species *O. glumaepatula* and clustering the cultivated species *japonica* and *indica* along with their direct progenitor species, *O. rufipogon* and *O. nivara*, respectively.(Khush 1997; Park et al. 2003) Moreover, the independent clustering of *O. sativa* ssp. *japonica* and *O. sativa* ssp. *indica* with closer affinity to their wild varieties, *O. rufipogon* and *O. nivara*, by NBS genes strongly supports diphyletic evolution of the Asian rice species(He et al. 2011; Park et al. 2003; Second 1982) and challenges the monophyletic origin or single domestication of the cultivated Asian rice proposed by researchers on the basis of a single domestication gene, *sh4* (non-shattering gene),(Vaughan et al. 2008) and a set of fragments from selected chromosomes, 8, 10, and 12.(Molina et al. 2011) Further, a separate cluster of the Australian *O. meridionalis* with the African *O. barthii* demonstrated the distinctness of the Australian AA genome wild rice from the Asian wild and cultivated rice species.(Waters et al. 2012) This is also consistent with the observations made earlier regarding the *O. meridionalis* genome.(Zhang et al. 2014)

Orthologous, Paralogous, and Species-Specific NBS Genes

Orthologous pairs were detected between predicted NBS-encoding genes at 40% identity over 70% of protein length, while 75% identity threshold was used to predict the paralogous pairs. Out of the predicted 2688 NBS-encoding genes, 166 genes (6.18%) were species-specific while the major proportion (93.82%) had orthologs in one or more *Oryza* species (Fig. 4, Table 3). A maximum number of orthologous genes were found in *O. barthii* (297) followed by *O. sativa* ssp. *indica* (296), *O. nivara* (291), and *O. rufipogon* (284) indicating that the progenitor wild species of cultivated rice were the most abundant compared to the other wild relatives. The highest number of orthologous pairs was found between *O. barthii-O. sativa* ssp. *indica* (1247) followed by the pairs of *O. rufipogon-O. sativa* ssp. *indica* (1220), *O. rufipogon-O. barthii* (1195), *O. rufipogon-O. nivara* (1165), and *O. sativa* ssp. *indica-O. nivara* (1155). These results revealed the robust orthology between the African and Asian progenitor species, though they diverged millions of years ago.(Zhang et al. 2014; Wambugu et al. 2015) Further, these observations clearly indicated that *O. sativa* ssp. *indica* still retain advantage over its counterpart *japonica* and *O. glaberrima* by keeping its reservoir of NBS genes rich. Since the number of NBS genes in *O. meridionalis* was extremely low, as expected, all combinations of the lowest number of orthologous pairs involved *O. meridionalis* exemplified

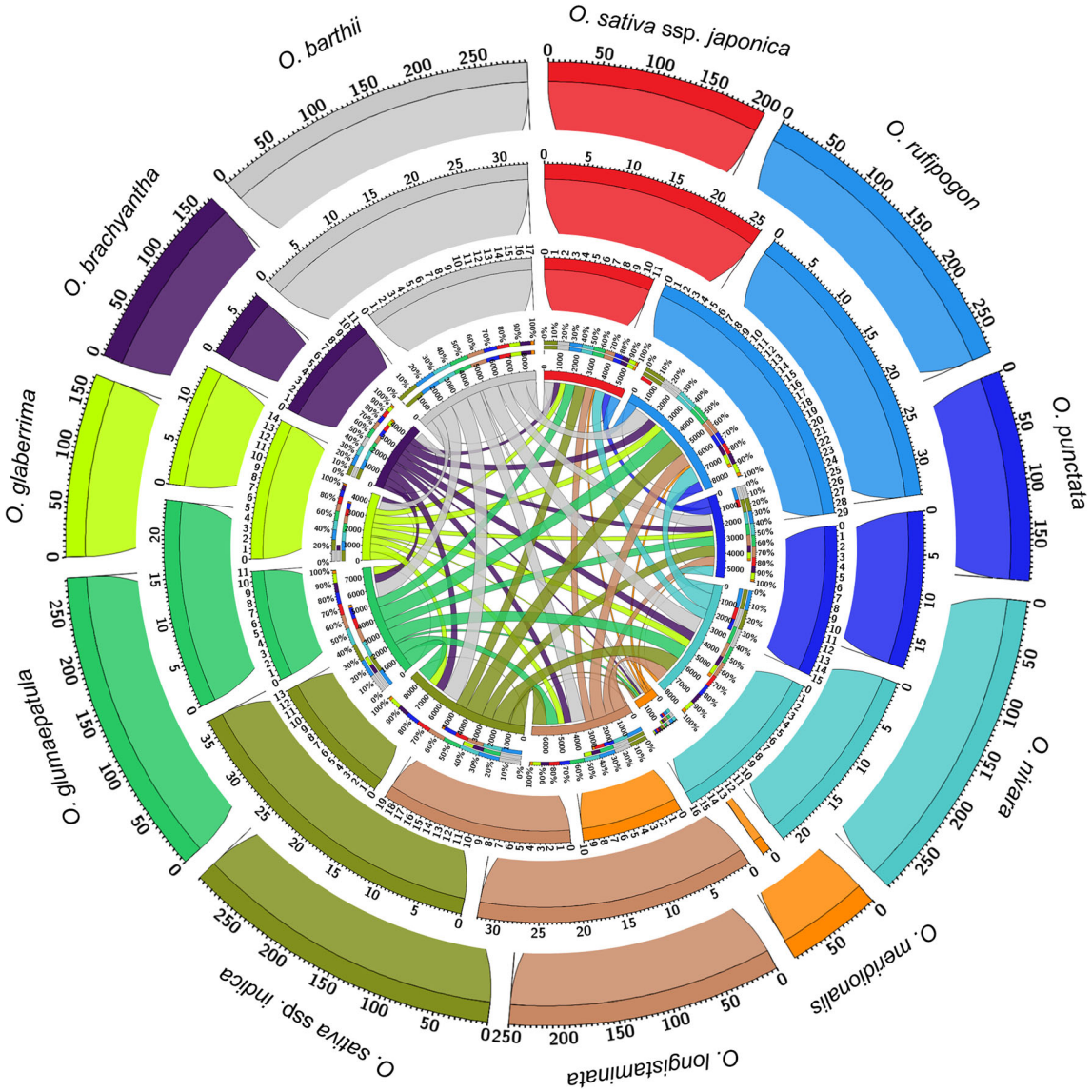

Fig. 4 A four-layered circos diagram with the outermost ring showing the number of genes having orthologs, followed by the number of genes with paralogs, the number of species-specific genes, and the number of orthologous and paralogous pairs between predicted NBS genes, moving from outermost to innermost ring

by *O. meridionalis-O. glaberrima* (91), *O. meridionalis-O. punctata* (113), *O. meridionalis-O. sativa* ssp. *japonica* (123), and *O. meridionalis-O. brachyantha* (129) (Fig. 4, Table S5). Thus, *O. meridionalis* had weak orthology not only with African and Asian cultivated rice but also with BB and FF genome *Oryza* species.

Species-specific gene is an important criterion to identify novel genes. *O. rufipogon* (29), followed by other wild species, *O. longistaminata* (19), *O. barthii* (17), *O. nivara* (16), and *O. punctata* (15), had the maximum number of species-specific NBS genes (Fig. 5). While there were 29 species-specific genes in the wild species, *O. rufipogon*, the cultivated species *O. sativa* ssp. *indica* and *O. sativa* ssp. *japonica* had just 13 and 11 unique genes as the genetic diversity declined due to domestication. This also suggested that for disease resistance, the best resources for novel NBS genes are *O. rufipogon*, *O. longistaminata*, *O. barthii*, *and O. nivara*. Interestingly, the South American species *O. glumaepatula* had the highest percentage of orthologous genes with other species (96.25% of predicted NBS genes, Table 3) with just 3.75% of species-specific NBS genes (Fig. 5).

A total of 248 NBS genes were predicted as paralogous genes at 75% identity over 50% length. The number of

Table 3 Number of paralogous and orthologous NBS genes between 11 *Oryza* species

Oryza sp.	Paralogous genes[1]			Orthologous genes[5] (genes having orthologs in rest of the species)	Species-specific genes[6]
	Low stringency[2] (~40)	Medium stringency[3] (~6 MYA)	High stringency[4] (~2 MYA)		
O. sativa ssp. *japonica*	26	8	2	202	11
O. sativa ssp. *indica*	37	10	6	296	13
O. rufipogon	34	6	0	284	29
O. punctata	18	6	2	192	15
O. nivara	22	6	0	291	16
O. meridionalis	0	0	0	87	10
O. longistaminata	32	11	5	250	19
O. glumaepatula	24	4	2	282	11
O. glaberrima	14	0	0	168	14
O. brachyantha	8	2	0	173	11
O. barthii	33	9	4	297	17
Total	248	62	21	2522	166

[1] Bidirectional hits with bit score ≥ 100 and *E* value ≤ $1e^{-20}$ and at least 75% identity

[2] At least 75% length of smaller gene is identical to ≥ 50% length of the larger gene of a paralog pair

[3] At least 85% length of smaller gene is identical to ≥ 70% length of the larger gene of a paralog pair

[4] At least 95% length of smaller gene is identical to ≥ 80% length of the larger gene of a paralog pair

[5] Bidirectional hits with bit score ≥ 100 and *E* value ≤ $1e^{-20}$ with at least 40% identity over 70% protein length

[6] Genes with no orthologs in other species

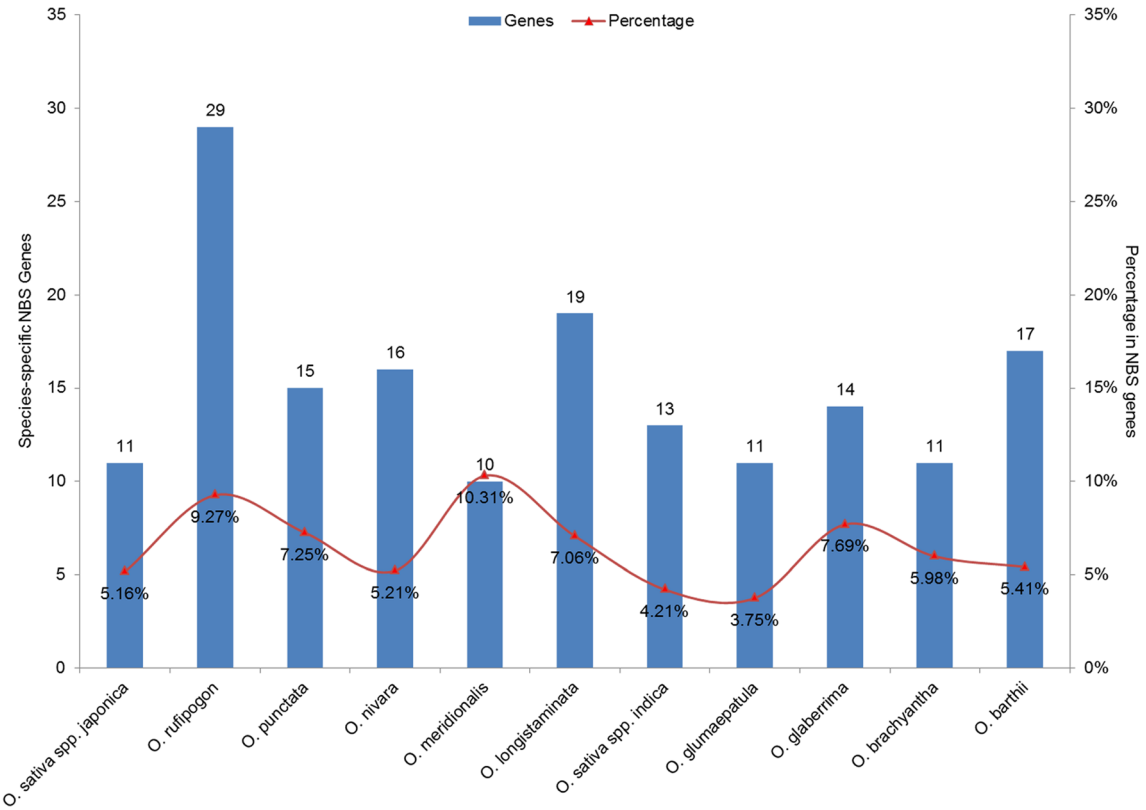

Fig. 5 Number and percentage of species-specific genes among the predicted NBS genes

paralogous NBS genes interestingly was the highest in *O. sativa* ssp. *indica* (37) suggesting continuous evolution of this species despite being cultivated, followed by 34, 33, and 32 in *O. rufipogon*, *O. barthii*, and *O. longistaminata*, respectively. These paralogous genes were categorized into three levels of duplications as more ancient, ancient, and recent, based on their stringency levels of identity percentages from low to high. Distribution pattern of paralogous genes for 11 *Oryza* species was the same at all three levels of low, medium, and high stringency. No paralogous genes in *O. meridionalis* at all three levels of stringency, in *O. glaberrima* at medium and high stringency levels, and in *O. rufipogon*, *O. nivara*, and *O. brachyantha* at high level of stringency suggested that divergence of the Australian and Asian AA genome rice predates the divergence of cultivated African AA genome (*O. glaberrima*) from Asian AA genome rice followed by the divergence of the wild Asian AA genome (*O. rufipogon* and *O. nivara*) (Table 3). *O. brachyantha* is the only species for which the predicted paralogous genes were less than the number of species-specific genes.

We observed that the number of predicted NBS-encoding genes was directly proportional to the number of orthologous and paralogous NBS genes and decreased from AA genome species to BB to FF with the exceptions of *O. glaberrima* and *O. meridionalis*. Similarly, the number of species-specific genes was proportional to the number of predicted NBS-encoding genes and orthologous and paralogous NBS genes except for *O. sativa* ssp. *indica* and *O. glumaepatula*.

Chromosomal Locations of Species-Specific Genes

A total of 166 NBS genes were identified across the 11 *Oryza* species. The chromosomal locations for these genes were detected by a BLAST search-based approach and visualized by MapChart. We observed that these genes were not evenly distributed among 12 chromosomes with chromosome 3 having just one gene (from *O. punctata*), while chromosomes 8 and 11 had a relatively high density of genes across all the 11 *Oryza* species (Fig. 6, Excel Data Sheet 2). Out of 166 genes, 66 genes were located in 26 gene clusters (i.e., two neighboring genes having chromosomal distance < 200 kb),(Holub 2001) comprising of two to five genes per clusters. There were two to five such clusters present on each of the chromosomes, except chromosomes 3, 5, 7, and 10. We carried out a comparison between species-specific NBS-encoding genes present in clusters and those not present in clusters to understand whether there are any structural or functional (presence/absence of motifs) differences between the two classes. We chose chromosome 11 for this analysis as this is the one with a maximum number of such genes. Eighteen genes were present in clusters and the same number as isolated ones on this chromosome. MSA as well as MEME was performed for these 36 genes which revealed no differences either in alignment or motif distribution between the two sets. The cultivated rice *O. sativa* ssp. *indica* had locus-specific NBS gene clusters with either *O. nivara* or *O. rufipogon*, the two wild progenitor Asian rice species. Significantly, no other species pairs showed a similar pattern of distribution of NBS genes on different chromosomes. This again demonstrated the direct descendance of the *indica* species from the wild ancestors, *O. nivara* or *O. rufipogon*. The 29 species-specific genes in *O. rufipogon* were distributed on eight chromosomes other than chromosomes 3, 5, 9, and 10. Since *O. rufipogon* had the highest number of species-specific genes, the similarity in physical clustering could be a boon in making crosses for introgression of the novel genes in *indica*.

Validation of Orthologous and Species-Specific Genes Through PCR

To validate the presence of orthologous (common) and species-specific genes, we randomly chose 12 genes with two of them being common to all species and two species-specific genes each from five different species, namely, *O. sativa* ssp. *japonica*, *O. sativa* ssp. *indica*, *O. rufipogon*, *O. nivara*, and *O. barthii*, and designed primers to enable their amplification (Table S6). Care was taken not to design the primer from the conserved NBS region. Orthologous genes were found to be amplified in all five species without exception (Fig. 7). Nivara-specific genes were found to be highly species-specific. While niv_1 amplification was absolutely species-specific, niv_2 gave amplification in one of the samples of *O. rufipogon* and *japonica*. The species specificity of *O. nivara* was observed even with *O. barthii*-specific genes, as for one of the genes, it showed a different product size. Another *O. barthii*-specific gene was found to be absent in *indica*. Overall, this demonstrated that all the predicted NBS genes are present in the genome with no false positives. Considering the established progenitor-descendent relationships of the cultivated species with rufipogon and nivara, *O. barthii* was the distinct genome used in the validation study. We did observe comparatively higher species specificity in *O. barthii* with different product sizes (for two genes) or no amplification (for four genes) in eight of the other species-specific primers tested. However, we did not observe strict species-specific amplification. The cross amplifications observed, despite using stringent criteria for determination of species-specific genes, could be due to the following reasons: (1) The primers targeted a small portion of the gene, while for prediction, the entire amino acid length

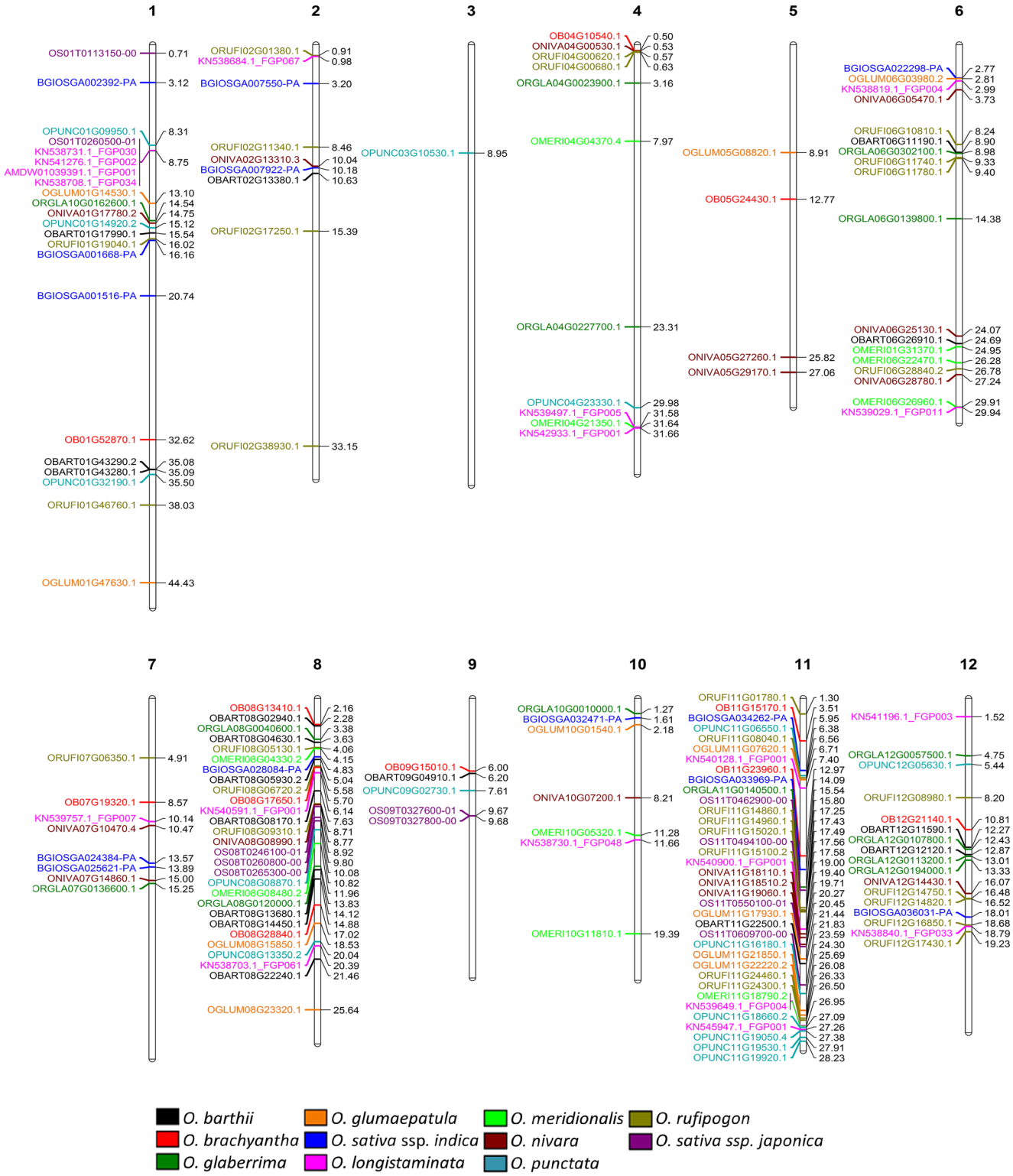

Fig. 6 Chromosomal location of predicted species-specific NBS genes in 11 *Oryza* species

was compared and those with < 40% identity over 70% length were named as species-specific. (2) Cryptic structural variations were present, especially in NBS domains. (3) The available sequence information did not represent intra-specific variation. This shows that the species-specific genes identified in the present study can be compared at sequence level across a panel of genotypes from the 11 species for their further validation and utility.

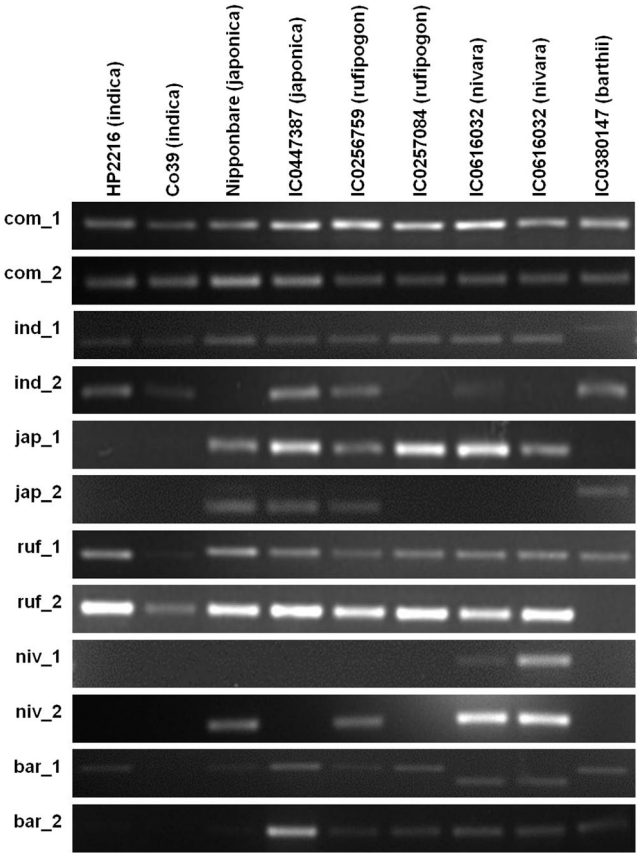

Fig. 7 Validation of orthologous and species-specific genes (com, common orthologous genes; species-specific genes: ind, indica; jap, japonica; ruf, rufipogon; niv, nivara; bar, barthii)

Conclusions

A total of 2688 NBS-encoding genes were identified across 11 *Oryza* genomes. We observed reduction in the number of NBS-encoding genes with evolutionary divergence from the AA genome species (314 to 269 genes) to BB genome (207 genes) and to FF genome (184 genes). The progenitors of the cultivated species, namely, *O. barthii* and *O. rufipogon*, were the richest species for NBS-encoding genes. In terms of proportion, *O. longistaminata* had the highest NBS-encoding genes (0.85%) while *O. meridionalis* had the least (0.22%). Chromosomes 11, 12, 8, and 1 harbored the maximum number of NBS genes in all the species. Nearly 94% of NBS-encoding genes had orthologs in one or the other *Oryza* species, while only 6% of genes were species-specific. African and Asian progenitor species were found to retain robust orthology whereas *O. meridionalis* showed weak orthology with African and Asian cultivated rice and with BB and FF genome *Oryza* species. Moreover, the South American AA genome species, *O. glumaepatula*, was distinctly separated from the rest of the AA genome species belonging to Asia and Africa. Thus, the phylogenetic relationship as revealed by the NBS-encoding genes was different from the established literature. The present NBS-encoding gene analyses not only provided a broader understanding of their structural and evolutionary features but established as a useful resource in resistance breeding.

Author Contributions AS and AM: conceived and designed the experiments; HR and NG: developed the pipelines and analyzed data; KA and VK: PCR validation; HR and AS: finalized tables and figures; DM, KC, and AR: provided the bioinformatic support; VD: supplied plant material; AM, AS, and TS: drafted and finalized the manuscript; and all the authors read and approved the manuscript.

Funding Information This research received financial support from the Centre for Agricultural Bioinformatics Scheme (CABin) funded by the Indian Council of Agricultural Research (ICAR), New Delhi, India.

Compliance with Ethical Standards

Conflict of Interest The authors declare that they have no conflict of interest.

References

Altschul S, Gish W, Miller W, Myers E, Lipman D (1990) Basic local alignment search tool. J Mol Biol 215(3):403–410

Ameline-Torregrosa C, Wang BB, O'Bleness MS et al (2008) Identification and characterization of NBS–LRR genes in the model plant *Medicago truncatula*. Plant Physiol 146:5–21

Andolfo G, Jupe F, Witek K, Etherington GJ, Ercolano MR, Jones JD (2014) Defining the full tomato NB-LRR resistance gene repertoire using genomic and cDNA RenSeq. BMC Plant Biol 14:120. https://doi.org/10.1186/1471-2229-14-120

Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, Madden TL (2009) BLAST+: architecture and applications. BMC Bioinf 10:421

Cheng X, Jiang H, Zhao Y, Qian Y, Zhu S, Cheng B (2009) A genomic analysis of disease-resistance genes encoding nucleotide binding sites in *Sorghum bicolor*. Genet Mol Biol 33(2):292–297. https://doi.org/10.1590/S1415-47572010005000036

Das A, Soubam D, Singh PK, Thakur S, Singh NK, Sharma TR (2012) A novel blast resistance gene, Pi54rh cloned from wild species of rice, *Oryza rhizomatis* confers broad spectrum resistance to *Magnaporthe oryzae*. *Funct Integr Genomics* 12:215–228. https://doi.org/10.1007/s10142-012-0284-1

Devanna NB, Vijayan J, Sharma TR (2014) The Blast resistance gene Pi54of cloned from *Oryza officinalis* interacts with Avr-Pi54 through its novel non-LRR domains. PLoS One 9(8):e104840. https://doi.org/10.1371/journal.pone.0104840

Guo YL, Fitz J, Schneeberger K, Ossowski S, Cao J, Weigel D (2011) Genome-wide comparison of nucleotide-binding site-leucine-rich repeat-encoding genes in *Arabidopsis*. Plant Physiol 157:757–769. https://doi.org/10.1104/pp.111.181990

He Z, Zhai W, Wen H, Tang T, Wang Y, Lu X, Greenberg AJ, Hudson RR, Wu CI, Shi S (2011) Two evolutionary histories in the genome of rice: the roles of domestication genes. PLoS Genet 7:e1002100. https://doi.org/10.1371/journal.pgen.1002100

Heidel AJ, Lawal HM, Felder M et al (2011) Phylogeny-wide analysis of social amoeba genomes highlights ancient origins for complex intercellular communication. *Genome Res* 21:1882–1891. https://doi.org/10.1101/gr.121137.111

Holub EB (2001) The arms race is ancient history in *Arabidopsis*, the wildflower. Nat Rev Genet 2:516–527

Hulsen T, Huynen MA, J-de V, Groenen PMA (2006) Benchmarking ortholog identification methods using functional genomics data. Genome Biol 7:R31. https://doi.org/10.1186/gb-2006-7-4-r31

Jaillon O, Aury JM, Noel B et al (2001) The grapevine genome sequence suggests ancestral hexaploidization in major angiosperm phyla. Nature 449:463–467

Jones P, Binns D, Chang HY, Fraser M, Li W, McAnulla C, McWilliam H, Maslen J, Mitchell A, Nuka G, Pesseat S, Quinn AF, Sangrador-Vegas A, Scheremetjew M, Yong SY, Lopez R, Hunter S (2014) InterProScan 5: genome-scale protein function classification. Bioinformatics 30(9):1236–1240. https://doi.org/10.1093/bioinformatics/btu031

Joshi RK, Nayak S (2013) Perspectives of genomic diversification and molecular recombination towards R-gene evolution in plants. Physiol Mol Biol Plants 19:1–9

Khush GS (1997) Origin, dispersal, cultivation and variation of rice. Plant Mol Biol 35:24–34

Larkin MA, Blackshields G, Brown NP, Chenna R, McGettigan PA, McWilliam H, Valentin F, Wallace IM, Wilm A, Lopez R, Thompson JD, Gibson TJ, Higgins DG (2007) ClustalW and ClustalX version 2.0. Bioinformatics 23:2947–2948. https://doi.org/10.1093/bioinformatics/btm404.

Li J, Ding J, Zhang W, Zhang Y, Tang P, Chen JQ, Tian D, Yang S (2010) Unique evolutionary pattern of numbers of gramineous NBS-LRR genes. Mol Gen Genomics 283:427–438. https://doi.org/10.1007/s00438-010-0527-6

Liu W, Liu J, Triplett L, Leach JE, Wang GL (2014) Novel insights into rice innate immunity against bacterial and fungal pathogens. Annu Rev Phytopathol 52:213–241. https://doi.org/10.1146/annurev-phyto-102313-045926

Lozano R, Hamblin MT, Prochnik S, Jannink JL (2015) Identification and distribution of the NBS-LRR gene family in the Cassava genome. BMC Genomics 16:360. https://doi.org/10.1186/s12864-015-1554-9

Lukasik E, Takken FL (2009) STANDing strong, resistance proteins instigators of plant defence. Curr Opin Plant Biol 12:427–436. https://doi.org/10.1016/j.pbi.2009.03.001

Marchler-Bauer A, Lu S, Anderson JB, Chitsaz F, Derbyshire MK, DeWeese-Scott C, Fong JH, Geer LY, Geer RC, Gonzales NR, Gwadz M, Hurwitz DI, Jackson JD, Ke Z, Lanczycki CJ, Lu F, Marchler GH, Mullokandov M, Omelchenko MV, Robertson CL, Song JS, Thanki N, Yamashita RA, Zhang D, Zhang N, Zheng C, Bryant SH (2011) CDD: a conserved domain database for the functional annotation of proteins. Nucleic Acids Res 39(Database issue): D225–D229

Meyers BC, Kozik A, Griego A, Kuang H, Michelmore RW (2003) Genome-wide analysis of NBS-LRR-encoding genes in *Arabidopsis*. Plant Cell 15:809–834. https://doi.org/10.1105/tpc.009308

Molina J, Sikora M, Garud N, Flowers JM, Rubinstein S, Reynolds A, Huang P, Jackson S, Schaal BA, Bustamante CD, Boyko AR, Purugganan MD (2011) Molecular evidence for a single evolutionary origin of domesticated rice. Proc Natl Acad Sci 108:8351–8356

Nandety RS, Caplan JL, Cavanaugh K, Perroud B, Wroblewski T, Michelmore RW, Meyers BC (2013) The role of TIR-NBS and TIR-X proteins in plant basal defence responses. Plant Physiol 162:1459–1472. https://doi.org/10.1104/pp.113.219162

Nguyen BD, Brar DS, Bui BC, Nguyen TV, Pham LN, Nguyen HT (2013) Identification and mapping of the QTL for aluminum tolerance introgressed from new source, *Oryza rufipogon* Griff. into indica rice, (*Oryza sativa* L.). Theor Appl Genet 106:583–593

Park KC, Kim NH, Cho YS, Kang KH, Lee JK, Kim NS (2003) Genetic variations of AA genome *Oryza* species measured by MITE-AFLP. Theor Appl Genet 107:203–209

Porter BW, Paidi M, Ming R, Alam M, Nishijima WT, Zhu YJ (2009) Genome-wide analysis of *Carica papaya* reveals a small NBS resistance gene family. Mol Gen Genomics 281:609–626. https://doi.org/10.1007/s00438-009-0434-x

Rahman ML, Jiang W, Chu SH, Qiao Y, Ham TH, Woo MO, Lee J, Khanam MS, Chin JH, Jeung JU, Brar DS, Jena KK, Koh HJ (2009) High-resolution mapping of two rice brown plant hopper resistance genes, Bph20(t) and Bph21(t), originating from *Oryza minuta*. Theor Appl Genet 119:1237–1246

Ram T, Majumder ND, Krishnaveni D, Ansari MM (2007) Rice variety Dhanarasi, an example of improving yield potential and disease resistance by introgressing gene(s) from wild species (*O. rufipogon*). Curr Sci 92:987–992

Rawal HC, Singh NK, Sharma TR (2013) Conservation, divergence and genome wide distribution of PAL and POX A gene families in plants. Int J Genomics 13:1–10. https://doi.org/10.1155/2013/678969

Sanchez PL, Wing RA, Brar DS. 2014 The wild relative of Rice: genomes and genomics, in Genetics and genomics of rice, ed. Zhang Q. and Wing R.A. (Plant genetics and genomics: crops and models 5, Springer), 9–25. DOI https://doi.org/10.1007/978-1-4614-7903-1_2

Second G (1982) Origin of the genetic diversity of cultivated rice (*Oryza* spp.): study of the polymorphism scored at 40 isozyme loci. Jpn J Genet 57:25–57

Seo E, Kim S, Yeom SI, Choi D (2016) Genome-wide comparative analyses reveal the dynamic evolution of nucleotide-binding leucine-rich repeat gene family among Solanaceae plants. Front Plant Sci 7:1205. https://doi.org/10.3389/fpls.2016.01205

Sharma T, Rai A, Gupta S, Vijayan J, Devanna B, Ray S (2012) Rice blast management through host-plant resistance: retrospect and prospects. Agric Res 1(1):37–52. https://doi.org/10.1007/s40003-011-0003-5

Singh S, Chand S, Singh NK, Sharma TR (2015) Genome-wide distribution, organisation and functional characterization of disease resistance and defence response genes across rice species. PLoS One 10(4):e0125964. https://doi.org/10.1371/journal.pone.0125964

Tanweer FA, Rafii MY, Sijam K, Rahim HA, Ahmed F, Latif MA (2015) Current advance methods for the identification of blast resistance genes in rice. C R Biol 338:321–334. https://doi.org/10.1016/j.crvi.2015.03.001

Tuskan GA, DiFazio S, Jansson S, Bohlmann J, Grigoriev I, Hellsten U, Putnam N, Ralph S, Rombauts S, Salamov A, Schein J, Sterck L, Aerts A, Bhalerao RR, Bhalerao RP, Blaudez D, Boerjan W, Brun A, Brunner A, Busov V, Campbell M, Carlson J, Chalot M, Chapman J, Chen GL, Cooper D, Coutinho PM, Couturier J, Covert S, Cronk Q, Cunningham R, Davis J, Degroeve S, Dejardin A, dePamphilis C, Detter J, Dirks B, Dubchak I, Duplessis S, Ehlting J, Ellis B, Gendler K, Goodstein D, Gribskov M, Grimwood J, Groover A, Gunter L, Hamberger B, Heinze B, Helariutta Y, Henrissat B, Holligan D, Holt R, Huang W, Islam-Faridi N, Jones S, Jones-Rhoades M, Jorgensen R, Joshi C, Kangasjarvi J, Karlsson J, Kelleher C, Kirkpatrick R, Kirst M, Kohler A, Kalluri U, Larimer F, Leebens-Mack J, Leple JC, Locascio P, Lou Y, Lucas S, Martin F, Montanini B, Napoli C, Nelson DR, Nelson C, Nieminen K, Nilsson O, Pereda V, Peter G, Philippe R, Pilate G, Poliakov A, Razumovskaya J, Richardson P, Rinaldi C, Ritland K, Rouze P, Ryaboy D, Schmutz J, Schrader J, Segerman B, Shin H, Siddiqui A, Sterky F, Terry A, Tsai CJ, Uberbacher E, Unneberg P, Vahala J, Wall K, Wessler S, Yang G, Yin T, Douglas C, Marra M, Sandberg G, van de Peer Y, Rokhsar D (2006) The genome of black cottonwood, *Populus trichocarpa* (Torr. & Gray). Science 313:1596–1604

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–R227

Van Zee JP, Schlueter JA, Schlueter S, Dixon P, Sierra CA, Hill CA (2016) Paralog analyses reveal gene duplication events and genes under positive selection in *Ixodes scapularis* and other ixodid ticks. BMC Genomics 17:241. https://doi.org/10.1186/s12864-015-2350-2

Vaughan DA (1994) The wild relatives of rice: a genetic resources handbook. International Rice Research Institute, Manila, p 137

Vaughan DA, Lu BR, Tomooka N (2008) The evolving story of rice evolution. Plant Sci 174:394–408

Voorrips RE (2002) MapChart: software for the graphical presentation of linkage maps and QTLs. J Hered 93:77–78. https://doi.org/10.1093/jhered/93.1.77

Wambugu PW, Brozynska M, Furtado A, Waters DL, Henry RJ (2015) Relationships of wild and domesticated rices (*Oryza* AA genome species) based upon whole chloroplast genome sequences. Sci Rep 5:13957. https://doi.org/10.1038/srep13957

Waters DL, Nock CJ, Ishikawa R, Rice N, Henry RJ (2012) Chloroplast genome sequence confirms distinctness of Australian and Asian wild rice. *Ecol Evol* 2:211–217. https://doi.org/10.1002/ece3.66

Yang S, Zhang X, Yue JX, Tian D, Chen JQ (2008) Recent duplications dominate NBS-encoding gene expansion in two woody species. Mol Gen Genomics 280:187–198. https://doi.org/10.1007/s00438-008-0355-0

Yue JX, Meyers BC, Chen JQ, Tian DC, Yang SH (2012) Tracing the origin and evolutionary history of plant nucleotide-binding site-leucine-rich repeat (NBS-LRR) genes. New Phytol 193:1049–1063. https://doi.org/10.1111/j.1469-8137.2011.04006.x

Zhang QJ, Zhu T, Xia EH, Shi C, Liu YL, Zhang Y, Liu Y, Jiang WK, Zhao YJ, Mao SY, Zhang LP, Huang H, Jiao JY, Xu PZ, Yao QY, Zeng FC, Yang LL, Gao J, Tao DY, Wang YJ, Bennetzen JL, Gao LZ (2014) Rapid diversification of five *Oryza* AA genomes associated with rice adaptation. *Proc Natl Acad Sci USA* 111:E4954–E4962. https://doi.org/10.1073/pnas.1418307111