

RESEARCH ARTICLE

Open Access



# Allele mining unlocks the identification of RYMV resistance genes and alleles in African cultivated rice

Hélène Pidon<sup>1,2</sup>, Sophie Chéron<sup>1</sup>, Alain Ghesquière<sup>1</sup> and Laurence Albar<sup>1\*</sup> 

## Abstract

**Background:** Rice yellow mottle virus (RYMV) is a major rice pathogen in Africa. Three resistance genes, i.e. *RYMV1*, *RYMV2* and *RYMV3*, have been previously described. *RYMV1* encodes the translation initiation factor eIF(iso)4G1 and the best candidate genes for *RYMV2* and *RYMV3* encode a homolog of an *Arabidopsis* nucleoporin (CPR5) and a nucleotide-binding domain and leucine-rich repeat containing domain (NLR) protein, respectively. High resistance is very uncommon in Asian cultivated rice (*Oryza sativa*), with only two highly resistant accessions identified so far, but it is more frequent in African cultivated rice (*Oryza glaberrima*).

**Results:** Here we report the findings of a resistance survey in a reference collection of 268 *O. glaberrima* accessions. A total of 40 resistant accessions were found, thus confirming the high frequency of resistance to RYMV in this species. We analysed the variability of resistance genes or candidate genes in this collection based on high-depth Illumina data or Sanger sequencing. Alleles previously shown to be associated with resistance were observed in 31 resistant accessions but not in any susceptible ones. Five original alleles with a frameshift or untimely stop codon in the candidate gene for *RYMV2* were also identified in resistant accessions. A genetic analysis revealed that these alleles, as well as T-DNA insertions in the candidate gene, were responsible of RYMV resistance. All 40 resistant accessions were ultimately linked to a validated or candidate resistance allele at one of the three resistance genes to RYMV.

**Conclusion:** This study demonstrated that the *RYMV2* resistance gene is homologous to the *Arabidopsis* CPR5 gene and revealed five new resistance alleles at this locus. It also confirmed the close association between resistance and an amino-acid substitution in the leucine-rich repeat of the NLR candidate for *RYMV3*. We also provide an extensive overview of the genetic diversity of resistance to RYMV in the *O. glaberrima* species, while underlining the contrasted pattern of diversity between *O. glaberrima* and *O. sativa* for this trait. The different resistance genes and alleles will be instrumental in breeding varieties with sustainable field resistance to RYMV.

**Keywords:** Rice, *Oryza glaberrima*, RYMV, Resistance gene, CPR5

\* Correspondence: [laurence.albar@ird.fr](mailto:laurence.albar@ird.fr)

<sup>1</sup>DIADÉ, Univ. Montpellier, IRD, Montpellier, France

Full list of author information is available at the end of the article



© The Author(s). 2020 **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated in a credit line to the data.

## Background

*Oryza sativa*, domesticated in Asia, is cultivated in almost all rice-growing areas worldwide. However, an independent rice domestication process occurred in Africa, which gave rise to the cultivated species *Oryza glaberrima* [1, 2]. The more productive *O. sativa* species was introduced in East Africa more than 1000 years ago and in West Africa in the sixteenth century, and has progressively supplanted *O. glaberrima*. Breeding initiatives over the last 60 years have essentially concerned *O. sativa* varieties and have further widened the gap in yield potential between varieties of the two species. Nonetheless, *O. glaberrima* has specific traits of interest and adaptation to local stresses, such as drought, iron toxicity, infertile soils and weed competition [3, 4]. This rich source of gene diversity is of substantial breeding interest to increase rice yield in a setting of global warming and reduced inputs. *O. glaberrima* was thus introduced in breeding programs [5, 6] leading for instance to the New Rice for Africa (NERICA) varieties, that resulted from *O. sativa* × *O. glaberrima* inter-specific crosses and were successfully disseminated in the 2000s [7, 8].

The rice yellow mottle virus (RYMV) is endemic to Africa and responsible for significant rice crop losses in irrigated or lowland areas [9]. High resistance, defined as the lack of symptoms and of virus detection in the systemic leaves using serological methods, appears to be very uncommon in *O. sativa*, with only two highly resistant varieties identified so far [10, 11], whereas 31 highly resistant *O. glaberrima* accessions have been reported [12, 13]. Moreover, while the two *O. sativa* resistant varieties share the same allele of the *RYMV1* resistance gene, which encodes a translation initiation factor [14] and seems to restrict cell-to-cell virus movement [15], at least three different *RYMV1* resistance alleles evolved independently in *O. glaberrima* [12]. These results suggest that *O. glaberrima* diversity for this trait would be particularly useful for rice breeding.

Two additional resistance genes, i.e. *RYMV2* and *RYMV3*, have been mapped in *O. glaberrima* species. *RYMV2*-mediated resistance is associated with a 1 bp deletion, leading to a null allele of an homolog of the *Arabidopsis constitutive expression of pathogenesis related protein-5 (CPR5)* nucleoporin gene in both a bi-parental mapping population and a collection of *O. glaberrima* accessions [13]. In *Arabidopsis*, the *CPR5* nucleoporin gene is involved in the regulation of defense mechanisms and senescence [16, 17]. Considering the sequence homology and in line with previous studies [13, 18], the candidate gene for *RYMV2* is hereafter referred to as *CPR5-1*, although its nucleoporin role and involvement in defense mechanisms has yet to be documented in rice. More recently, a gene of the nucleotide binding domain and leucine-rich repeat gene (NLR) superfamily was

pinpointed as the best candidate for the *RYMV3* dominant resistance gene [19]. This gene is hereafter referred to as *NLR<sub>RYMV3</sub>*. Resistance is associated with a single amino-acid substitution in the leucine-rich repeat (LRR) domain of the protein, which is known to be involved in the pathogen recognition specificity.

Here we describe the diversity of RYMV resistance genes or candidates in one of the most documented *O. glaberrima* collections, which covers the geographical distribution of the species and includes 165 fully sequenced accessions [20, 21]. We also validated the candidate gene for *RYMV2* using natural variants identified in *O. glaberrima* diversity and *O. sativa* T-DNA mutants.

## Results

### Screening for resistance to RYMV in a collection of *O. glaberrima* accessions

Thiemele et al. [12] and Orjuela et al. [13] screened 120 accessions of the *O. glaberrima* collection described in Orjuela et al. [20] for resistance to RYMV and found 31 highly resistant accessions. In the present study, these 31 accessions and 148 additional ones from the same collection were phenotyped for resistance by double antibody sandwich enzyme-linked immunosorbent assay (DAS-ELISA) performed on samples from systemic

**Table 1** Phenotyping of *O. glaberrima* accessions for RYMV resistance

Accessions	Ratio infected/inoculated	Infection rate (%)
Susceptible controls	22/24	92
Tog5681	0/8	0
Tog7291	2/12	17
Og26	0/18	0
Og111	0/12	0
Og133	6/17	35
Og183	2/18	11
Og213	4/13	31
Og256	4/18	22
Og406	1/16	6
Og423	0/4	0
Og447	2/16	13
Og452	0/16	0
Og491	5/23	23
Og498	2/18	11

Only accessions identified as resistant in this study are listed in this table. Resistance was evaluated based on ELISA tests performed on individual plants after mechanical inoculation with the BF1 isolate of RYMV. The first screening was performed on a set of four plants and confirmed, when seeds were available, in additional screening experiments. Tog5681 and Tog7291, carrying resistance alleles on *RYMV1* and *RYMV2* genes, respectively, were used as resistance controls and Og82, Og431 and CG14 were used as susceptible controls

leaves of four plants per accession. The same virus isolate as that reported in Thiemele et al. [12] and Orjuela et al. [13] was used. Of the 31 accessions previously reported as being resistant, we confirmed the resistance of 28, while three were susceptible, presumably because of between seed stocks heterogeneity. All four plants of most of the 148 newly tested accessions were clearly susceptible. However, highly resistant plants were observed in 12 accessions for which the high resistance phenotype was confirmed in additional plants when seeds were available (Table 1; Additional file 1: Table S1). For eight of those, up to a third of the plants multiplied the virus, suggesting incomplete penetrance of resistance or possible resistance-breaking events, as previously reported [18, 19, 22]. However, the rate of susceptible vs. resistant plants was not significantly different than observed on the Tog7291 accession carrying the *RYMV2* major gene (Fisher exact test,  $p > 0.05$ ) and these accessions were thus considered resistant. Finally, the accessions identified as being resistant in this study were: Og26, Og111, Og133, Og183, Og213, Og256, Og406, Og423, Og447, Og452, Og491 and Og498 (Table 1; Additional file 1: Table S2). A total of 40 accessions out of 268 were therefore highly resistant to the BF1 isolate of RYMV.

#### Allele mining in RYMV resistance genes or candidates

Among resistant *O. glaberrima* accessions, previous results indicated that 12 have a resistance allele on the *RYMV1* gene [12], 7 have an allele associated with *RYMV2*-mediated resistance on the *CPR5-1* gene [13], 1 has a resistance allele on *RYMV3*, for which *NLR<sub>RYMV3</sub>* is a candidate [19], while the Tog5672 accession carries a resistance allele on both *RYMV1* and *RYMV3* [19]. The genes or alleles responsible for resistance in the 19 remaining accessions were unknown.

Polymorphisms in the *RYMV1* gene, and in the *RYMV2* and *RYMV3* candidate genes, the *CPR5-1* gene and *NLR<sub>RYMV3</sub>*, respectively, were analyzed in 174 accessions, including the 40 resistant accessions and 134 susceptible ones. For 165 accessions, the full genome sequence was available [1, 21, 23]. For nine additional resistant accessions, the partial or complete sequence of the target genes were obtained from Thiemele et al. [12] or by Sanger sequencing of polymerase chain reaction (PCR) fragments.

#### Allele mining in RYMV1

A total of ten single nucleotide polymorphisms (SNPs) or small insertions/deletions (indels), defining nine different haplotypes, were detected in the *RYMV1* gene (Additional file 1: Table S3). The three most frequent haplotypes at the nucleotidic level represented 83% of the accessions and corresponded to the protein variant of the CG14 accession, while the others were detected in

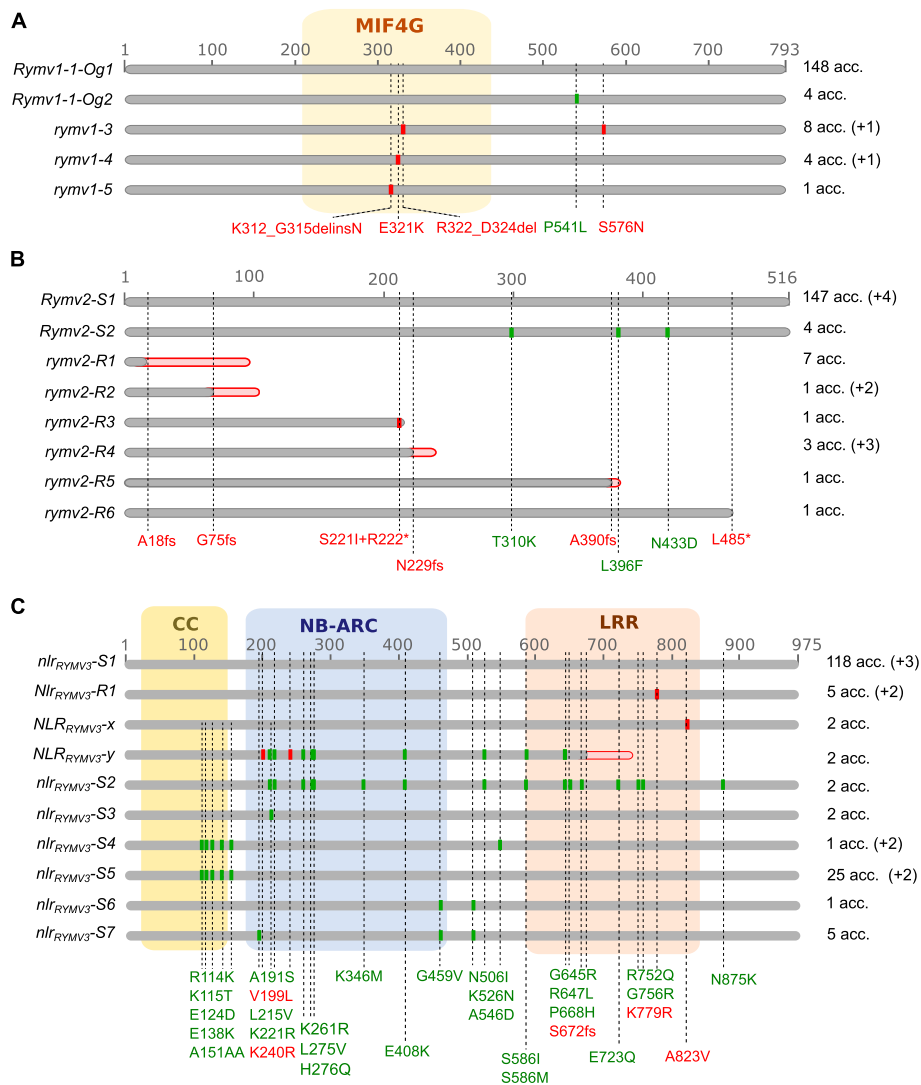
less than 5% of the accessions. Five mutations – three SNPs and two indels – were located in the exons and all resulted in amino-acid changes (Fig. 1a). One of them that caused a single amino-acid substitution (P541L) was present in susceptible accessions. The others, which were previously described as characterizing *rymv1-3* (R322\_D324del, S576N), *rymv1-4* (E321K) and *rymv1-5* (K312\_G315delinsN) resistance alleles [12, 14], were associated with resistance in the full collection. Fifteen resistant accessions carried those resistance alleles (Table 2; Additional file 1: Table S2), including two accessions identified as resistant in this study, Og208 and Og423, which carried alleles *rymv1-3* and *rymv1-4*, respectively.

#### Allele mining in CPR5-1

In the *CPR5-1* gene, 12 polymorphisms were detected at the nucleotidic level based on genomic data from Cubry et al. [21] (Additional file 1: Table S4). However, the filters used in this analysis hampered detection of the 1 bp-deletion on codon 17 that characterized the allele of the resistant Tog7291 accession [13] because it is located in an artificially created SNP-cluster, probably due to a GCC rich region [24]. Nevertheless, all accessions of the collection had been previously genotyped at this position based on a CAPS marker [13]. Moreover, the deletion was confirmed by manual curation of the read alignment data (BAM file) of the Tog7291 accession. The 13 polymorphisms in the *CPR5-1* gene defined eight haplotypes at the nucleotide level and eight protein variants (Fig. 1b; Additional file 1: Table S4). The CG14 reference haplotype was observed in 89% of the accessions. Six haplotypes were characterized by frameshifts (A18fs, G75fs, N229fs, A390fs) or an untimely stop codon (R222\*, L485\*), leading to truncated forms of the protein, while conserving from 3 to 93% of the protein sequence. Interestingly, these haplotypes concerned 19 accessions that were all highly resistant to RYMV (Table 2; Additional file 1: Table S2), with the most frequent being the Tog7291 haplotype that was previously described in seven accessions [13], while the others were less frequent haplotypes that were found in one to three accessions. Finally, four RYMV-susceptible accessions (Og186, Og426, Og459 and Og89) shared the same haplotype characterized by three SNPs in the introns and three SNPs causing amino-acid substitutions (T310K, L396F, N433D).

#### Allele mining in NLR<sub>RYMV3</sub>

The variability in *NLR<sub>RYMV3</sub>*, with 66 polymorphisms at the nucleotidic level, was far greater than the variability observed in *RYMV1* and *CPR5-1* (Additional file 1: Table S5). Yet, the polymorphisms identified in the first intron were probably underestimated because of the



**Fig. 1** Variants of resistance genes or candidates at the protein level. For *RYMV1* (a), the allele names cited in Albar et al. [14] and Thiemele et al. [12] are used, but an additional protein variant observed in susceptible accessions was given the name “*Rymv1-1-Og2*”, and for greater clarity the allele named “*Rymv1-1-Og*” in [12] was referred to as “*Rymv1-1-Og1*”. For the *RYMV2* (b) and *RYMV3* (c) candidate genes, the different alleles were named according to their association (R) or not (S) with RYMV resistance. The CG14 allele was the reference allele. Polymorphisms associated with resistance are indicated in red, whereas those which are not are in green. Important conserved domains are indicated as colored frames. The number of accessions carrying each allele is indicated on the right of the figure, with a distinction between accessions from the set of 165 fully sequenced accessions (without brackets), and accessions from the nine additional resistant accessions (in brackets). The total number of accessions sometimes differed between genes due to missing data, resulting in undefined alleles

marked differences between the Nipponbare sequence used as mapping reference and the CG14 sequence, which probably hampered correct mapping and extensive SNP calling in this region. Eleven haplotypes were detected, with the CG14 haplotype being found in 71.5% of the accessions. Forty-nine mutations were located in exons, including 35 that were non-synonymous (Fig. 1c). These mutations defined ten protein variants, three of which were specific to resistant accessions (Table 2; Additional file 1: Table S5); two displayed a single amino-acid substitution compared to the reference allele

(K779R described in Pidón et al. [19], and A823V), while the third one showed a frameshift in the LRR domain (S672fs) and 11 amino-acid substitutions. The K779R mutation was observed in the two accessions – Tog5307 and Tog5672 – known to carry a resistance allele of *RYMV3* [19], and five that displayed a resistant phenotype but did not carry resistance specific alleles on *RYMV1* or *CPR5-1*, which suggested that their resistance may be associated with the K779R mutation. Conversely, accessions carrying the A823V mutation (Og20, Og36) also had known resistance alleles of *RYMV1*

**Table 2** List of the resistant accessions and their alleles on the resistance genes or candidates

ID in [20]	Alternate ID	Allele on RYMV1	Allele on RYMV2 candidate	Allele on RYMV3 candidate
Og8	Tog5321	<i>rymv1-3</i>		
Og13	Tog5418	<i>rymv1-3</i>		
Og20	Tog5438	<i>rymv1-4</i>		<i>NLR<sub>RYMV3</sub>-X</i>
Og24	Tog5463	<i>rymv1-4</i>		
Og26	Tog5474			<i>Nlr<sub>RYMV3</sub>-R1</i>
Og28	Tog5486	<i>rymv1-3</i>		
Og36	Tog5556	<i>rymv1-3</i>		<i>NLR<sub>RYMV3</sub>-X</i>
Og44	Tog5672	<i>rymv1-4</i>		<i>Nlr<sub>RYMV3</sub>-R1</i>
Og46	Tog5674	<i>rymv1-5</i>		
Og49	Tog5681	<i>rymv1-3</i>		
Og87	Tog6220, IRGC112577		<i>rymv2-R1</i>	<i>NLR<sub>RYMV3</sub>-Y</i>
Og103	Tog6698		<i>rymv2-R1</i>	<i>NLR<sub>RYMV3</sub>-Y</i>
Og111	Tog5286			<i>Nlr<sub>RYMV3</sub>-R1</i>
Og126	Tog5747			<i>Nlr<sub>RYMV3</sub>-R1</i>
Og133	RAM 131		<i>rymv2-R2</i>	
Og150	Tog7206, IG38		<i>rymv2-R1</i>	
Og151	Tog14367		<i>rymv2-R4</i>	
Og152	Tog13943		<i>rymv2-R4</i>	
Og153	Tog13709		<i>rymv2-R4</i>	
Og160	Tog7235, IRGC103549		<i>rymv2-R1</i>	
Og161	Tog10434		<i>rymv2-R1</i>	
Og163	Tog12086			<i>Nlr<sub>RYMV3</sub>-R1</i>
Og164	Tog12160			<i>Nlr<sub>RYMV3</sub>-R1</i>
Og165	Tog12188	<i>rymv1-4</i>		
Og166	Tog12249	<i>rymv1-3</i>		
Og172	Tog7291		<i>rymv2-R1</i>	
Og183	Tog7456		<i>rymv2-R3</i>	
Og186	Tog8049	<i>rymv1-3</i>		
Og207	Tog12386	<i>rymv1-3</i>		
Og208	Tog12387	<i>rymv1-3</i>		
Og213	Tog12401		<i>rymv2-R4</i>	
Og256	Tog14361		<i>rymv2-R6</i>	
Og406	Tog7197		<i>rymv2-R4</i>	
Og411	Tog7202, IG35		<i>rymv2-R1</i>	
Og423	IRGC104018	<i>rymv1-4</i>		
Og447	1LG104		<i>rymv2-R2</i>	
Og452	3LG117B		<i>rymv2-R2</i>	
Og491	OG1		<i>rymv2-R5</i>	
Og498	SG329		<i>rymv2-R4</i>	
Og6_2	Tog5307			<i>Nlr<sub>RYMV3</sub>-R1</i>

Only alleles associated with resistance are indicated



(*rymv1-4* and *rymv1-3*, respectively), and accessions carrying the S672fs mutation (Og87, Og103) had a *CPR5-1* allele associated with resistance.

Ultimately, all resistant accessions described in the *O. glaberrima* collection carried an allele associated with resistance in at least one of the three analyzed genes (Table 2).

### Comparison with *O. sativa*

Moreover, we looked for polymorphisms at *RYMV1*, *CPR5-1* and *NLR<sub>RYMV3</sub>* in *O. sativa* based on the SNP-Seek database [25], which pools genotyping data from the 3000 Rice Genomes Project [26]. Seventeen non-synonymous mutations were identified in *RYMV1*. They resulted in amino-acid substitutions or small deletions, but only three occurred in the middle domain of the eukaryotic initiation factor 4G (MIF4G), where all mutations conferring resistance to RYMV were located (Additional file 1: Table S6). One of them (A303D) was present only in the few *O. glaberrima* accessions included in the 3000 Rice Genomes Project, as well as in all accessions from our *O. glaberrima* collection. This mutation was therefore considered to be specific to *O. glaberrima* and not associated with resistance to RYMV. The two others, i.e. K352R and P395S, were detected in four and ten accessions, respectively. While located in the MIF4G domain, they did not occur in the 15 amino-acid region which was mutated in the resistance alleles described so far, but instead were detected at least 28 amino-acids downstream. Twenty-three non-synonymous mutations were detected in the *CPR5-1* gene (Additional file 1: Table S6). However, none of them led to an untimely stop codon or frameshift. Similarly to what we observed in our *O. glaberrima* dataset, *O. sativa* presented high variability at the *NLR<sub>RYMV3</sub>*

locus, with 112 non-synonymous mutations (Additional file 1: Table S6). Eight mutations were detected in 10.4% of the accessions and resulted in stop codons or frameshifts. The 104 others were in frame mutations, leading to amino-acid substitutions or single amino acid insertions or deletions in the protein. Interestingly, three *O. sativa* spp. *indica* accessions carried the K779R mutation associated with RYMV resistance in *O. glaberrima*. These accessions shared a very specific haplotype, with 27 additional uncommon non-synonymous mutations that differentiated them from both other *O. sativa* and *O. glaberrima* accessions.

### Loss-of-function mutations in the *CPR5-1* gene confer resistance to RYMV

A genetic analysis was performed to check the association between the truncated *CPR5-1* forms identified in *O. glaberrima* and RYMV resistance. Resistant Og256, Og213, Og491, Og133 and Og183 accessions, representing the different truncated forms of *CPR5-1*, were crossed with a susceptible *O. glaberrima* accession (Og82 or Og431) and with the resistant Tog7291 accession, whose resistance is controlled by *RYMV2* [13]. F2 seeds were obtained for all combinations except (Og183 x Tog7291), and at least 45 F2 plants per population were phenotyped for RYMV resistance. The resistance segregations noted in all populations developed with the susceptible Og82 or Og431 accessions were in agreement with a 1R:3S segregation ratio (Table 3), indicating monogenic and recessive control of resistance. Genotyping on the *CPR5-1* gene was performed on a total of 281 plants based on Sanger sequencing for the Og133-derived population and cleaved amplified polymorphic sequence (CAPS) or derived cleaved amplified polymorphic sequence (dCAPS) markers for all the other

**Table 3** Cosegregation of RYMV resistance and allelic state on *CPR5-1*

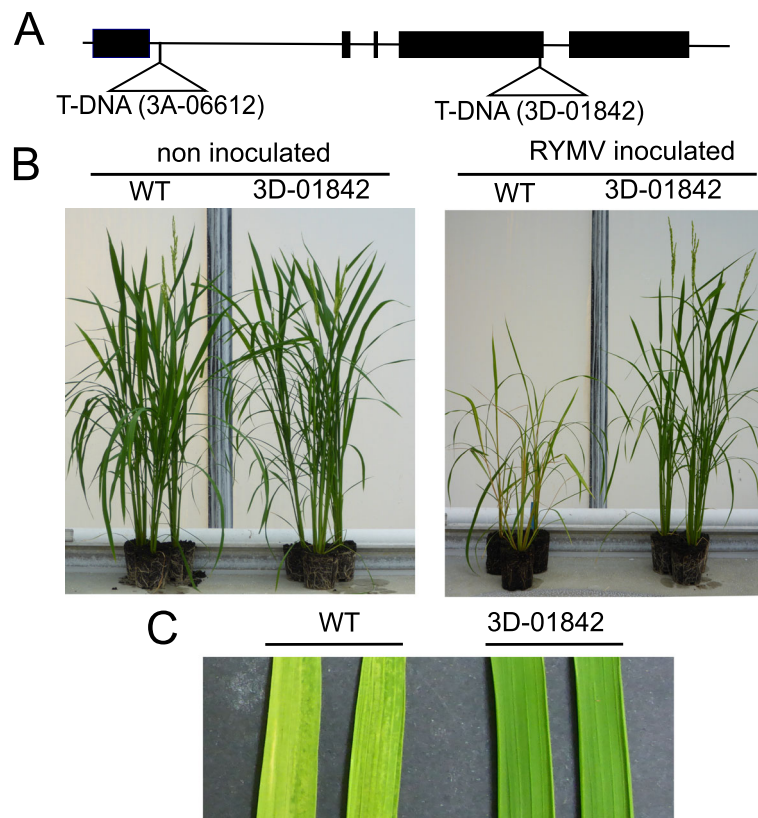
F2 population	Phenotype and genotype			
	Total	<i>rymv2-Rx</i>	Htz	<i>Rymv2-S1</i>
Og133 ( <i>rymv2-R2</i> ) x Og431 ( <i>Rymv2-S1</i> )	11 R, 50S ( $p = 0,23$ )	10 R	22 S	3 S
Og133 ( <i>rymv2-R2</i> ) x Tog7291 ( <i>rymv2-R1</i> )	55 R, 1S			
Og183 ( <i>rymv2-R3</i> ) x Og82 ( <i>Rymv2-S1</i> )	14 R, 36 S ( $p = 0,62$ )	14 R	26 S	10 S
Og213 ( <i>rymv2-R4</i> ) x Og82 ( <i>Rymv2-S1</i> )	13 R, 42 S ( $p = 1$ )	13 R	31 S	11 S
Og213 ( <i>rymv2-R4</i> ) x Tog7291 ( <i>rymv2-R1</i> )	100 R			
Og491 ( <i>rymv2-R5</i> ) x Og431 ( <i>Rymv2-S1</i> )	24 R, 47 S ( $p = 0,10$ )	21 R	3 R, 30 S	17 S
Og491 ( <i>rymv2-R5</i> ) x Tog7291 ( <i>rymv2-R1</i> )	56 R			
Og256 ( <i>rymv2-R6</i> ) x Og82 ( <i>Rymv2-S1</i> )	18 R, 52 S ( $p = 0,89$ )	18 R, 2S	37 S	13 S
Og256 ( <i>rymv2-R6</i> ) x Tog7291 ( <i>rymv2-R1</i> )	45 R			

F2 plants were evaluated for RYMV resistance based on symptom observations. The phenotype is indicated with R for resistant plants and S for susceptible ones. For populations derived from crosses with Og82 and Og431 susceptible accessions, the observed phenotypic segregation was compared to a 1R/3S segregation ratio, based on a binomial exact test. For these populations, genotyping on the *CPR5-1* gene was generally performed on all plants with CAPS or dCAPS markers, except for the (Og133 x Og431) population for which genotyping was based on Sanger sequencing and only performed on a subset of 35 plants. The genotype is indicated as "*rymv2-Rx*" for plants homozygous for alleles *rymv2-R2* to *-R6*, "*Rymv1-S1*" for plants homozygous for the *Rymv1-S1* allele, and Htz for heterozygous plants. F2 plants derived from crosses with Tog7291 were not genotyped

populations. Most of the plants homozygous for a truncated form of the protein were resistant (76 out of 79), while most of the others (200 out of 202) were susceptible, showing a close association between the *CPR5-1* allelic state and RYMV resistance. We hypothesized that the five plants that did not fit this pattern were misclassified, presumably because of lack of inoculation or resistance breakdown. Besides, 256 F2 plants from populations developed with the resistant Tog7291 accession were resistant (Table 3), while a single one was susceptible. These results demonstrated that the different truncated forms of *CPR5-1* were resistance alleles of the *RYMV2* recessive resistance gene.

In addition, *O. sativa* lines mutated in the *CPR5-1* gene were characterized. T-DNA insertional mutant lines tagged in the *CPR5-1* gene were identified by searching the flanking sequence database [27] of the mutant library developed by Jeon et al. [28] and Jeong et al. [29]. Two independent T-DNA insertions in the *CPR5-1* gene were confirmed by sequencing the T-DNA flanking regions. In the 3D-01842 line, T-DNA was inserted 1975 bp downstream of the ATG, in the fourth exon; in

the 3A-06612 line, T-DNA was inserted 315 bp downstream of the ATG in the first intron (Fig. 2a). Phenotyping of these mutants was performed on a minimum of 12 plants homozygous for the insertion. Ten weeks post-sowing, 3D-01842 and 3A-06612 non-inoculated mutants did not show any visible differences in plant morphology or development compared to the wild-type controls (Fig. 2b). The mutants inoculated with RYMV did not show any symptoms or growth reduction compared to the non-inoculated controls, while wild-type plants expressed very clear yellowing and mottling symptoms, a marked growth reduction or even growth arrest (Fig. 2b and c). In addition, contrary to wild-type plants, mutants did not accumulate the virus according to the ELISA test findings. A total of 117 pseudo-T3 or F3 plants which segregated for one or another T-DNA insertion were analyzed. Except for one plant, a perfect co-segregation was observed between resistance and T-DNA insertions at the homozygous state (Table 4). This indicated that, in both *O. sativa* and *O. glaberrima*, altered forms of *CPR5-1* lead to RYMV resistance.



**Fig. 2** Characterization of T-DNA mutant lines 3D-01842 and 3A-06612. **a** Structure of the *CPR5-1* gene and positions of T-DNA, inserted 315 bp downstream of the ATG in the 3A-06612 line and 1975 bp downstream of the ATG in the 3D-0184 line. Exons are represented as black boxes and introns as black lines. **b** Phenotype of wild-type controls and the 3D-0184 mutant on 10-week old non-inoculated plants, and 8 weeks after RYMV inoculation on inoculated plants, at the whole plant level. **c** Yellowing of leaves of wild-type controls and the 3D-0184 mutant 2 weeks post-inoculation with RYMV

**Table 4** Segregation of T-DNA and RYMV resistance in progenies

Mutant	WT	Htz	Mut
3D-01842	12 S	34 S	15 R
3A-06612	14 S	23 S	18 R, 1S

A pseudo-T3 progeny derived from the 3D-01842 mutant, and F3 progenies derived from the 3A-06612 mutant were analyzed. The phenotype is indicated with "R" for resistance and "S" for susceptibility. The genotype is indicated with "WT" for plants without the T-DNA insertion, "Mut" for plants homozygous for the T-DNA insertion and "Htz" for plants heterozygous at the T-DNA insertion site

## Discussion

The results of Orjuela et al. [13] strongly suggested that *CPR5-1* is the *RYMV2* gene, conferring resistance to RYMV. Here we validated this hypothesis using two independent T-DNA mutants in *O. sativa* and six different alleles leading to truncated forms of the protein in *O. glaberrima*. Although *Arabidopsis cpr5* mutants are known to be affected in biotic resistance [16], this is the first time that this gene has been described as a natural resistance gene in a crop species. In *Arabidopsis*, *CPR5* is a transmembrane nucleoporin involved in the membrane ring of the nuclear pore complex [17]. Loss of function mutations permeabilize the nuclear pore and mediate the activation of cell cycle transcription factors, leading to defense gene expression. Constitutive resistance to several pathogens is one of the resulting phenotypes, but the mutant shows additional deleterious developmental phenotypes, such as reduced size [16] and seed yield [30], which would be incompatible with breeding strategies for biotic resistance in crops. Some of the six *RYMV2* resistance alleles identified in *O. glaberrima* were very probably null alleles as stop or frame-shift mutations were found to occur far upstream, whereas the *rymv2-R6* allele retained 93% of the wild-type protein sequence. Whether the protein completely loses its cellular function or not remains to be investigated. However, based on the homology with *Arabidopsis* [17], even the *rymv2-R6* allele would lack at least one of the transmembrane domains. Unexpectedly, no obvious deleterious phenotype appeared to be associated with these mutations. In addition, the detection of several null alleles that have evolved independently and have been maintained hardly supports a strong deleterious effect of *CPR5-1* knock-out. This could be explained by the presence of two *Arabidopsis CPR5* homologs in rice [13] which may have partial functional redundancy. It is also possible that the functional homolog of *Arabidopsis CPR5* is *CPR5-2* and not *CPR5-1*. The cellular function of each copy will have to be further analyzed. However, our results suggest that the use of null or truncated *CPR5-1* alleles in rice breeding programs, either by introgression from *O. glaberrima* or by mutagenesis, would be an effective way to achieve RYMV

resistance. Similar mutations may provide resistance in other pathosystems and allele mining in species that harbor two homologs of *Arabidopsis CPR5*, such as other cereal species, may help uncover new pathways of pathogen resistance.

Contrary to what was observed for accessions carrying *RYMV1* or *RYMV3* resistance alleles, no accessions carrying a *RYMV2* resistance allele showed resistance in 100% of the plants screened. This may have resulted from incomplete penetrance of resistance, as previously described in other plant/virus interactions [31, 32], or resistance-breaking events. The high rate of resistance-breakdown reported by Pinel-Galzi et al. [18] on the Tog7291 accession carrying the *rymv2-R1* resistance allele suggests that resistance-breaking events is the most likely hypothesis. Indeed, they reported resistance-breaking rates of up to 96% depending on isolates, while other results reported on *RYMV1* [33] and *RYMV3* [19] suggested less frequent resistance-breakdown on those genes.

This study also revealed new resistance sources without a resistance allele at the *RYMV2* locus. The *RYMV1* locus has been the focus of extensive analysis in recent years [12, 14, 34]. This larger scale study revealed two additional accessions carrying known resistance alleles but did not uncover any new resistance alleles. On the *NLR<sub>RYMV3</sub>* gene, we identified five additional accessions showing the K779R amino-acid substitution in the LRR region that was proposed as being responsible for a high resistance phenotype [19]. These five accessions displayed high resistance to RYMV, which further strengthens the candidate status of the *NLR<sub>RYMV3</sub>* gene, and particularly the K779R mutation, but formal functional validation is still needed to confirm this. Two other haplotypes at *NLR<sub>RYMV3</sub>* were specific to resistant accessions but the corresponding accessions carried alleles on *RYMV1* and *RYMV2*, which would be sufficient to explain their high resistance level. Furthermore, the *NLR<sub>RYMV3</sub>-γ* sequence variant was characterized by a truncated LRR domain, suggesting a loss of function, which is not consistent with a gain of resistance. We think it is likely that those two sequence variants do not confer resistance to RYMV but further genetic analyses would be necessary to confirm this.

In contrast, we did not find any convincing candidate resistance alleles on *RYMV1* and *RYMV2* genes among accessions from the 3000 Rice Genomes Project [26], which mostly includes *O. sativa* accessions. At the *RYMV1* locus, two rare mutations were identified in the MIF4G domain and would require further analysis. However, based on the predicted 3D structure of the MIF4G domain [14], they occurred downstream of the α-helical hairpin that forms a protrusion where mutations known to be responsible for high resistance are



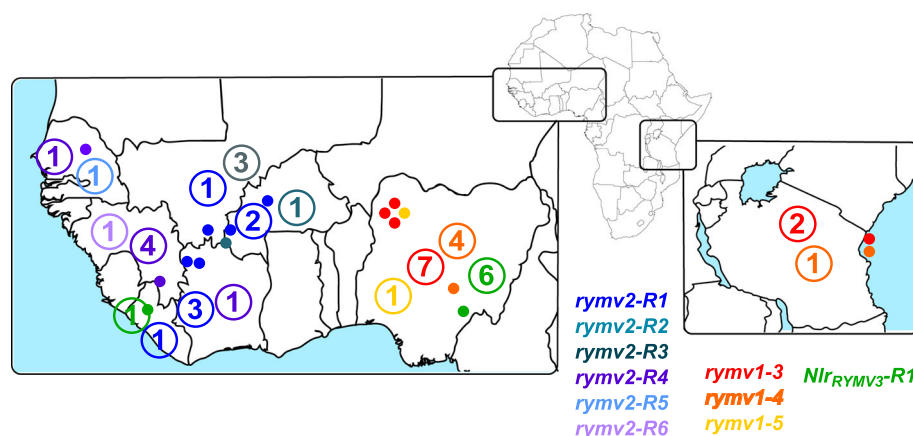
located. We therefore do not consider these mutations as likely candidates for resistance. High variability was observed at the  $NLR_{RYMV3}$  locus and a simple sequence analysis would not be sufficient to pinpoint mutations that may be involved in resistance. In particular, the K779R mutation, which is associated with resistance in *O. glaberrima*, has been detected in three *O. sativa* spp. *indica* accessions. However it is hard to speculate on their resistance, as these accessions were also characterized by 14 additional rare mutations in the LRR domain.

The probable absence of candidate resistance alleles on *RYMV1* and *RYMV2* within accessions of the 3000 Rice Genomes Project—mainly *O. sativa*, as mentioned above—is in agreement with the scant resistance observed in this species based on phenotypic screening. Indeed, only two accessions with a high level of resistance to RYMV, like those described in this study, have been reported in *O. sativa* [10, 11]. These two accessions, originating from East Africa, both carry the *rymv1-2* resistance allele. This result contrasted with the relatively high number of resistance alleles detected in *O. glaberrima*. Out of the 268 accessions of the collection used in this study, 40 highly resistant accessions were detected, which corresponded to approximately 15% of the collection. Yet this rate was probably overestimated because about ten accessions previously identified as resistant [12, 35] were deliberately included when the collection was set up [20]. The actual rate of resistant accessions in *O. glaberrima* is probably closer to 8%, which is the rate calculated on the basis of the 148 accessions newly evaluated in this study and for which we did not have any a priori knowledge. Still, this rate is very much higher than in *O. sativa*.

The diversity profiles on *RYMV1*, *RYMV2* and  $NLR_{RYMV3}$  genes were contrasted. First, we observed a

high number of mutations at the  $NLR_{RYMV3}$  gene, with 35 non-synonymous mutations detected in the *O. glaberrima* collection. Such high variability was expected and has been widely documented for the *NLR* gene family, which is known to be hypermutagenic and frequently under balancing or diversifying selection as a result of the arms race between plants and pathogens [36–41]. Secondly, *RYMV1* and *RYMV2* presented lower variability, with five and ten non-synonymous mutations detected, respectively, in accordance with their central role in plant cells. Indeed, *RYMV1* codes for eIF(iso)4G-1 [14], a translation initiation factor that is part of the cell translation machinery, while the *Arabidopsis* gene homologous to *RYMV2* codes for a component of the nuclear pore complex. These two genes are therefore assumed to be under conservative selection. Interestingly, three out of five non-synonymous mutations in *RYMV1* and six out of ten in *RYMV2* were directly involved in the resistance phenotype. In a similar gene/pathogen interaction, the results of Charron et al. [42] provided evidence of diversifying selection on the eIF4E locus that would at least partially be driven by potyvirus-induced selection pressure. As RYMV emerged quite recently, in the mid-nineteenth century [43, 44], there has not been a long co-evolution between the virus and *O. glaberrima* that could have explained the allelic diversity observed at the resistance loci. However, selection pressure on these loci may have been exerted by other viruses using these same plant factors.

The different resistance genes and alleles were positioned on the genetic diversity tree of the species proposed by Orjuela et al. [20] and on a map according to the geographical origin of the accessions. For all three resistance genes, accessions carrying the same resistance allele were generally showing a similar geographic origin



**Fig. 3** Geographical origin of accessions carrying resistance alleles on *RYMV1*, *RYMV2* and  $NLR_{RYMV3}$  genes. The geographical origins of the accessions were obtained from Cubry et al. [21]. Accessions for which GPS coordinates were available are represented by colored points. In each country, the total number of accessions carrying a specific allele (with or without GPS coordinates) is indicated as a number. The background map comes from [Freemove.org](https://www.freemove.org) and is available under the Creative Commons 0 license

(Fig. 3) and were clustered on the genetic diversity tree (Additional file 2: Figure S1), as expected since *O. glaberrima* has geographically-based population structuring [45]. More surprisingly, accessions with different *RYMV1* or *RYMV2* resistance alleles also appeared to be clustered, despite the independence of the mutations characterizing the alleles. Accessions with *RYMV2* resistance alleles were located west of the Benin-Niger axis, while accessions with *RYMV1* resistance alleles were located east of this axis. *RYMV3* apparently did not fit this distribution pattern, but the low number of accessions limited the scope of these findings. Several hypotheses may explain the observed *RYMV1* and *RYMV2* structuring. First, both eIF(iso)4G and CPR5-1 – if confirmed as a nucleoporin – are part of large protein complexes. Their variability may have been driven by the genetic structuring of other members of the same complexes. Besides, the environmental conditions may have led to a difference in selection pressure on *RYMV1* and *RYMV2*, or members of the complex in which they participate. Accessions with *RYMV2* resistance alleles mainly originated from regions of historically dense rice cultivation, while accessions with *RYMV1* resistance alleles originated from regions where rice was more sparsely cultivated, according to Portères [46]. This axis is also compatible with the separation of two distinct genetic groups of RYMV isolates [43, 44], one of which – the easternmost – includes hypervirulent isolates able to overcome most known resistance sources [33]. As underlined previously, the diversification of resistance genes under a selection pressure exerted by RYMV is quite improbable, but the opposite hypothesis has yet to be investigated, along with the impact of environmental conditions on both the virus and resistance gene diversity. Moreover, like *A. thaliana* *CPR5* [17], if *RYMV2* is a regulator of effector-triggered immunity and programmed cell death, it may confer resistance to several pathogens and could have evolved under selection pressure exerted by another pathogen. We would be unable to perform a more detailed population genetics analysis due to the limited number of resistant accessions available, but additional collections have been described [45, 47] and should now be investigated.

## Conclusions

Our results highlighted the allelic diversity in the three known resistance genes against RYMV. All 40 *O. glaberrima* accessions identified as being highly resistant in this study carried at least one of the confirmed or candidate resistance alleles on *RYMV1*, *RYMV2* and *NLR<sub>RYMV3</sub>* (Table 2). This suggests that we have probably identified most of the resistance genes that occur in *O. glaberrima*, even though wild relative species, such as *Oryza barthii*, may also contain original resistance

sources. Sound knowledge on resistance genes against RYMV and their diversity is thus now available, as well as a good assessment of the frequency and molecular determinants of resistance-breakdown in controlled conditions [18, 22, 33]. This knowledge provides an opportunity to design strategies of resistance gene deployment that will optimize resistance durability. Previous results suggest that all three genes are effective against a large spectrum of RYMV isolates. However, the high capacity of some virus isolates to evolve and overcome resistance prompts breeders to manage the use of resistance genes by pyramiding, variety mixtures or temporal rotation, depending on the local virus diversity. So far, a single *RYMV1* resistance allele originating from *O. sativa* accessions has been transferred into high-yielding varieties [48, 49] that are about to be deployed in the field. The use of additional resistance alleles or genes, and combinations, should be promoted to increase RYMV resistance sustainability.

## Methods

### Plant material

The *O. glaberrima* collection used in this study was described in Orjuela et al. [20]. This collection was jointly established by the French National Research Institute for Sustainable Development (IRD) and the Africa Rice Center and the accessions studied were selected for their current breeding impact and geographical distribution. All accessions are registered in the International Treaty on Plant Genetic Resources for Food and Agriculture. The primary origin of seeds is listed in Orjuela et al. [20] and seeds were multiplied in IRD greenhouses. Accessions identifiers (ID) of Orjuela et al. [20] are generally used, except for the CG14 reference accession or accessions previously characterized for RYMV resistance for which the previously used names were adopted [12, 13]. Correspondences between the different ID are given in Additional file 1: Table S2.

T-DNA mutant lines were obtained from the Pohang University of Science and Technology, Pohang, Korea [28, 29]. The 3D-01842 line was derived from the Hwayoung variety and the 3A-06612 line from the Dongjin variety. For the 3D-01842 line, a plant heterozygous for the T-DNA insertion in the *CPR5-1* gene was self-pollinated to develop pseudo-T2 and pseudo-T3 progenies that segregated for the insertion. As no plant heterozygous for the T-DNA insertion was available for the 3A-06612 line, a plant homozygous for the insertion was crossed with the Dongjin variety and F1 hybrids were selfed to successively derive F2, and F3 progenies that segregated for the T-DNA insertion.

### Resistance evaluation

Plants were grown in greenhouses and mechanically inoculated about 2 weeks after sowing with an RYMV

isolate originating from Burkina Faso (BF1). The resistance evaluations were based on symptom observation and confirmed when necessary with DAS-ELISA to estimate the virus content in systemic leaf samples harvested 2–3 weeks post-inoculation. Details on the inoculation and ELISA protocols were previously described in Pinel-Galzi et al. [50]. Samples were considered positive when optical density, minus the value of the negative control (buffer), are superior to 0,1.

### Genomic data

Genomic data on *O. glaberrima* accessions based on high coverage genomic sequencing (average 35X, range 20–55X) were obtained from Cubry et al. [21]. The IRGSP 1.0 Nipponbare sequence [51] had been used as reference for mapping and SNP calling. The polymorphism database [52, 53] included genomic data of 163 *O. glaberrima* and 83 *O. barthii* accessions. In a first step, the database was filtered for the 163 *O. glaberrima* accessions and for the ATG-stop codon regions of target genes. The ORGLA04G0147000.1, ORGLA01G0359000.1 and ORGLA11G0175800.1 gene models established on the CG14 accession [1], corresponding to Os04g42140.1, Os01g68970.1 and Os11g43700.1 gene models on the reference Nipponbare sequence, were considered for *RYMV1*, *CPR5-1* and *NLR<sub>RYMV3</sub>*, respectively. The target regions corresponded to positions 24,946,655–24,952,068 on chromosome 4 of the reference Nipponbare sequence for *RYMV1*, 40,071,092–40,073,727 on Nipponbare chromosome 1 for *CPR5-1* and 26,377,263–26,380,577 on Nipponbare chromosome 11 for *NLR<sub>RYMV3</sub>*. Only SNPs and indels that were polymorphic within *O. glaberrima* accessions were conserved. In a second step, polymorphisms were filtered with GATK 3.7 VariantFiltration [54] using the following filters:  $QUAL < 200$ ,  $MQ0 > 4$  &  $MQ0/DP > 0.1$ ,  $DP < 10$ , clusterSize 3 in clusterWindow-Size 10,  $DP > 20,000$ . SNPs with more than 10% missing data or heterozygous in more than 10% accessions were filtered out, and genotypes defined based on a single read were considered as missing data. Genomic data of the *O. glaberrima* CG14 [1] and Tog5681 accessions [23] were included in the dataset. The haplotype of the RYMV-susceptible CG14 accession was used as a reference to describe variants observed in the population.

Genomic data from the 3000 Rice Genomes Project [26] were retrieved from the Rice SNP-Seek database [25]. The database was filtered on the target regions indicated above for indels and non-synonymous SNPs from the base SNP set, which includes about 18 million SNPs. SNP effects were retrieved from the database, while indel effects were manually estimated.

The sequence variant nomenclature proposed by Den Dunnen et al. [55] was used to describe the mutations and their effects on CDS and proteins. Based on the

results described in this paper or previously [12, 14, 19], dominant alleles were indicated with the first letter in upper case and recessive ones with the first letter in lower case; when there was no preferred hypothesis, allele names were written with all letters in upper case. For the *RYMV2* and *RYMV3* candidate genes, the different alleles were named depending on their association (R) or not (S) with RYMV resistance.

### Sanger sequencing and molecular markers

PCR amplifications were performed on leaf extracts or DNA, as described in Orjuela et al. [13]. Primers were designed using Primer3 [56], except for the primers used for dCAPS markers, which were designed with the dCAPS Finder tool [57]. Primer sequences are provided in the Additional file 2: Figure S2. Partial or complete Sanger sequencing of *RYMV1*, *CPR5-1* and *NLR<sub>RYMV3</sub>* genes was performed on PCR amplification fragments and subcontracted to Genewiz (Takeley, UK).

CAPS and dCAPS markers were designed on the *CPR5-1* gene to genotype polymorphic loci identified in *O. glaberrima* accessions. Marker characteristics are described in the Additional file 2: Figure S2 and Table S7.

The T-DNA segregation analysis was based on the amplification of a T-DNA-specific fragment and a gene-specific fragment involving a common primer. Primer sequences and positions are indicated in the Additional file 2: Figure S2.

### Supplementary information

**Supplementary information** accompanies this paper at <https://doi.org/10.1186/s12870-020-02433-0>.

**Additional file 1: Table S1.** Example of DAS-ELISA results on *O. glaberrima* accessions. DAS-ELISA were performed on the systemic leaves of individual plants harvested 15 days after inoculation as described in [50]. OD is the optical density at 405 nm; ODcor represents OD minus the mean of negative (buffer) controls. Samples were considered positive when ODcor are superior to 0,1. Tog5681 and Tog7291 were included as resistant controls; CG14 and Og82 as susceptible ones. This table presents data acquired in a single experiment, on a subset of 8 plants per accession, on all the accessions identified as resistant in this study, except Og423. **Table S2.** ID, phenotype and genotype of accessions characterized for RYMV resistance. Resistance to RYMV was evaluated after mechanical inoculation of the BF1 isolate in this study or in previous studies [12, 13]. Alleles on resistance genes or candidates refer to the results presented in the Additional file 1: Table S3, Table S4, Table S5 or in previous studies [12]. **Table S3.** Genotype on the *RYMV1* resistance gene. Only positions where polymorphisms were detected in the *O. glaberrima* collection analyzed in Cubry et al. [21] were included. Nucleotide positions refer to the IRGSP1.0 reference sequence of the *O. sativa* Nipponbare accession [51] that was used as mapping reference. The effect of the mutations are based on the ORGLA04G0147000.1 gene model established on the *O. glaberrima* CG14 accession [1]. Mutations are described according to the nomenclature proposed by Den Dunnen et al. [55], except that synonymous mutations and mutations occurring in an intron are denoted “syn” and “intron”, respectively. Different variants at

the protein level were considered as different alleles. Names for resistance alleles were previously attributed by Albar et al. [14] and Thiemele et al. [12], but an additional protein variant observed in susceptible accessions was given the name “*Rymv1-1-Og2*”, and for greater clarity the allele named “*Rymv1-1-Og*” in [12] was referred to as “*Rymv1-1-Og1*”. **Table S4.** Genotype on the *CPR5-1* gene, candidate for *RYMV2*. Only positions were polymorphisms were detected in the *O. glaberrima* collection analyzed in Cubry et al. [21] were included. Nucleotide positions referred to the IRGSP1.0 reference sequence of the *O. sativa* Nipponbare accession [51] that was used as mapping reference. The effects of the mutations are based on the ORGLA01G0359000.1 gene model established on the *O. glaberrima* CG14 accession [1]. Mutations are described according to the nomenclature proposed by Den Dunnen et al. [55], except that synonymous mutations and mutations occurring in an intron are noted “syn” and “intron”, respectively. Different variants at the protein level were considered as different alleles. The allele names were chosen to distinguish protein variants associated or not with RYMV resistance. **Table S5.** Genotype on the *NLR<sub>RYMV3</sub>* gene, candidate for *RYMV3*. Only positions were polymorphisms were detected in the *O. glaberrima* collection analyzed in Cubry et al. [21] were included. Nucleotide positions refer to the IRGSP1.0 reference sequence of the *O. sativa* Nipponbare accession [51] that was used as mapping reference. The effects of the mutations are based on the ORGLA11G0175800.1 gene model established on the *O. glaberrima* CG14 accession [1]. Mutations are described according to the nomenclature proposed by Den Dunnen et al. [55], except that synonymous mutations and mutations occurring in an intron are noted “syn” and “intron”, respectively. Different variants at the protein level were considered as different alleles. The allele names were chosen to distinguish protein variants associated or not with RYMV resistance. **Table S6.** Diversity on RYMV resistance genes or candidates in accessions from the 3000 Rice Genomes Project [26]. Only non-synonymous SNPs from the base SNPs set are reported here. SNP effects were retrieved from the SNP-Seek database [25] and indels effects were evaluated manually. The effects of mutations on CDS and proteins are based on the Os04g42140.1 and Os01g68970.1 gene models established on the Nipponbare IRGSP1.0 sequence [51], for *RYMV1* and *CPR5-1*, respectively. For *NLR<sub>RYMV3</sub>* the CDS is based on the Os11g43700.1 gene model, except that the ATG codon was shifted from 180 nucleotides downstream of the original starting codon to best fit the corresponding CDS of the ORGLA11G0175800.1 gene model established on CG14 reference sequence. Effects on the CDS and protein were thus adapted. Frequency refers to the percentage of the alternate variant in the complete set of accessions. Mutations located in the PFAM domains MA3, MIF4G and LRR and in the HMM Panther hit LRR are indicated.

**Additional file 2: Figure S1.** Positions of accessions with resistance alleles of *RYMV1*, *RYMV2* and *NLR<sub>RYMV3</sub>* genes on the genetic diversity tree. Susceptible accessions are colored in dark grey and accessions not evaluated for resistance in light grey. Adapted from the genetic tree of Orjuela et al. [20]. **Figure S2.** Characteristics of primers and amplified fragments for markers or Sanger sequencing. Genes are represented as grey boxes for exons and grey lines for introns. Primers are represented as triangles and the numbers below the triangle refer to the corresponding sequences. (a, b, c) Blue traits represent fragments that were amplified and then sequenced with the primers colored in red. (c) Amplification fragments corresponding to the CAPS or dCAPS markers designed on the *CPR5-1* gene are represented as green traits. Additional information on these markers is provided in Additional file 2: Table S7. (d) Position of T-DNA insertions in the *CPR5-1* gene in lines 3A-06612 and 3D-01842 are indicated. The T-DNA-specific and gene-specific primers used for sequencing the T-DNA flanking site and genotyping for the presence/absence of insertions are indicated in brown and blue, respectively. **Table S7.** Characteristics of CAPS and dCAPS markers. Marker names indicate whether there are CAPS or dCAPS markers and which alleles of the *CPR5-1* gene they target. The bracketed number before the primer sequences refer to the reference of primers in Additional file 2: Figure S2. The size of the fragments expected in plants with the reference haplotype of CG14 (WT) or the alternate haplotypes (R) are indicated, except fragments below 30 bp that are uneasily detected by agarose electrophoresis. The CAPS-*CPR5-1*-R1 marker had already been described in Orjuela et al. [13].

## Abbreviations

CPR5: Constitutive expression of pathogenesis-related protein-5; DAS-ELISA: Double antibody sandwich-enzyme-linked immunosorbent assay; dCAPS: Derived cleaved amplified polymorphic sequence; ID: Identifier; Indel: Small insertion/deletion; MIF4G: Middle domain of eukaryotic initiation factor 4G; LRR: Leucine-rich repeat; NLR: Nucleotide-binding domain and leucine-rich repeat containing domain; PCR: Polymerase chain reaction; RYMV: Rice yellow mottle virus; SNP: Single nucleotide polymorphism

## Acknowledgements

We thank Christine Dubreuil-Tranchant and François Sabot for their help in the bioinformatics analysis, Gatean Maillot for technical help in T-DNA mutant characterization and Harold Chrestin for seed multiplication. We are also grateful to Denis Fargette, François Anthony, Mathias Lorieux and Nils Poulicard for helpful discussions. We thank Gynheung An for providing the T-DNA mutant lines.

## Authors' contributions

AG, HP and LA planned and designed the experiments, SC developed the populations, SC, HP and LA performed phenotyping, genotyping and sequencing, HP and LA analyzed the data and wrote the paper, AG and SC revised the paper. All authors have read and approved the manuscript.

## Funding

This work was financially supported by the CGIAR Research Program on rice agri-food systems (RICE, 2017–2022). The French *Ministère de l'Enseignement Supérieur et de la Recherche* provided a PhD grant to H. Pidón. Funding bodies did not participate in the design of the study and collection, analysis, and interpretation of data and in writing the manuscript.

## Availability of data and materials

The *O. glaberrima* genomic dataset analysed in the current study have been described in [21] and are available in the IRD Gigwa repository, <https://gigwa.ird.fr/gigwa/>. The sequences obtained by Sanger in this study are available in GenBank under reference accessions numbers MT348348 to MT348356 for *RYMV2*, MT348357 to MT348360 for *RYMV1*, and MT348361 to MT348369 for *NLR<sub>RYMV3</sub>*. All additional data generated in this study are included in the present article and its supplementary information files.

## Ethics approval and consent to participate

Not applicable.

## Consent for publication

Not applicable.

## Competing interests

The authors declare that they have no competing interests.

## Author details

<sup>1</sup>DIADÉ, Univ. Montpellier, IRD, Montpellier, France. <sup>2</sup>Present Address: Leibniz Institute of Plant Genetics and Crop Plant Research (IPK) Gatersleben, Seeland, Germany.

Received: 28 January 2020 Accepted: 7 May 2020

Published online: 19 May 2020

## References

- Wang M, Yu Y, Haberer G, Marri PR, Fan C, Goicoechea JL, et al. The genome sequence of African rice (*Oryza glaberrima*) and evidence for independent domestication. *Nat Genet.* 2014;46:982–8.
- Meyer RS, Choi JY, Sanches M, Plessis A, Flowers JM, Amas J, et al. Domestication history and geographical adaptation inferred from a SNP map of African rice. *Nat Genet.* 2016;48(9):1083–8.
- Linares OF. African rice (*Oryza glaberrima*): history and future potential. *Proc Natl Acad Sci U S A.* 2002;99(25):16360–5.
- Sarla N, Mallikarjuna Swamy BP. *Oryza glaberrima*: a source for the improvement of *Oryza sativa*. *Curr Sci.* 2005;89(6):955–63.
- Jones MP, Dingkuhn M, Aluko GK, Semon M. Interspecific *Oryza sativa* L. x *O. glaberrima* Steud. progenies in upland rice improvement. *Euphytica.* 1997;94(2):237–46.



6. Heuer S, Miezan KM, Sie M, Gaye S. Increasing biodiversity of irrigated rice in Africa by interspecific crossing of *Oryza glaberrima* (Steud.) x *O. sativa indica* (L.). Euphytica. 2003;132(1):31–40.
7. Kijima Y, Otsuka K, Sserunkuma D. Assessing the impact of NERICA on income and poverty in central and western Uganda. Agric Econ. 2008;38(3): 327–37.
8. Dibba L, Zeller M, Digne A. The impact of new Rice for Africa (NERICA) adoption on household food security and health in the Gambia. Food Secur. 2017;9(5):929–44.
9. Kouassi NKK, N'Guessan P, Albar L, Fauquet CM, Brugidou C. Distribution and characterization of rice yellow mottle virus : a threat to African farmers. Plant Dis. 2005;89(2):124–33.
10. Ndjondjop M-N, Albar L, Fargette D, Fauquet C, Ghesquière A. The genetic basis of high resistance to rice yellow mottle virus (RYMV) in cultivars of two cultivated rice species. Plant Dis. 1999;83(10):931–5.
11. Rakotomalala M, Pinel-Galzi A, Albar L, Ghesquière A, Rabenantoandro Y, Ramavololona P, et al. Resistance to rice yellow mottle virus in rice germplasm in Madagascar. Eur J Plant Pathol. 2008;122(2):277–86.
12. Thiémélé D, Boissard A, Ndjondjop M-N, Chéron S, Séré Y, Aké S, et al. Identification of a second major resistance gene to rice yellow mottle virus, RYMV2, in the African cultivated rice species, *O. glaberrima*. Theor Appl Genet. 2010;121(1):169–79.
13. Orjuela J, Thiémélé D, Kolade O, Chéron S, Ghesquière A, Albar L. A recessive resistance to rice yellow mottle virus is associated with a rice homolog of the *CPR5* gene, a regulator of active defense mechanisms. Mol Plant-Microbe Interact. 2013;26(12):1455–63.
14. Albar L, Bangratz-Reyser M, Hebrard E, Ndjondjop M-N, Jones M, Ghesquière A. Mutations in the eIF(iso)4G translation initiation factor confer high resistance of rice to rice yellow mottle virus. Plant J. 2006;47(3):417–26.
15. Ndjondjop MN, Brugidou C, Zang S, Fargette D, Ghesquière A, Fauquet C. High resistance to rice yellow mottle virus in two cultivated rice cultivars is correlated with failure of cell to cell movement. Physiol Mol Plant Pathol. 2001;59:309–16.
16. Bowling SA, Clarke JD, Liu Y, Klessig DF, Dong X. The *cpr5* mutant of *Arabidopsis* expresses both NPR1-dependent and NPR1-independent resistance. Plant Cell. 1997;9(9):1573–84.
17. Gu Y, Zebell SG, Liang Z, Wang S, Kang B-H, Dong X. Nuclear pore permeabilization is a convergent signaling event in effector-triggered immunity. Cell. 2016;166:1–13.
18. Pinel-Galzi A, Dubreuil-Tranchant C, Hébrard E, Mariac C, Ghesquière A, Albar L. Mutations in rice yellow mottle virus polyprotein P2a involved in RYMV2 gene resistance breakdown. Front Plant Sci. 2016;7:1–11.
19. Pidón H, Ghesquière A, Chéron S, Issaka S, Hébrard E, Sabot F, et al. Fine mapping of RYMV3: a new resistance gene to rice yellow mottle virus from *Oryza glaberrima*. Theor Appl Genet. 2017;130(0):807–18.
20. Orjuela J, Sabot F, Chéron S, Vigouroux Y, Adam H, Chrestin H, et al. An extensive analysis of the African rice genetic diversity through a global genotyping. Theor Appl Genet. 2014;127(10):2211–23.
21. Cubry P, Tranchant-Dubreuil C, Thuillet AC, Monat C, Ndjondjop M-N, Labadie K, et al. The rise and fall of African rice cultivation revealed by analysis of 246 new genomes. Curr Biol. 2018;28:2274–82.
22. Traoré O, Pinel-Galzi A, Issaka S, Poulicard N, Aribi J, Aké S, et al. The adaptation of rice yellow mottle virus to the eIF(iso)4G-mediated rice resistance. Virology. 2010;408(1):103–8.
23. Monat C, Pera B, Ndjondjop MN, Sow M, Tranchant-Dubreuil C, Bastianelli L, et al. *De novo* assemblies of three *Oryza glaberrima* accessions provide first insights about pan-genome of African rices. Genome Biol Evol. 2017;9(1):1–6.
24. Nakamura K, Oshima T, Morimoto T, Ikeda S, Yoshikawa H, Shiwa Y, et al. Sequence-specific error profile of Illumina sequencers. Nucleic Acids Res. 2011;39(13):e90.
25. Mansueto L, Fuentes RR, Borja FN, Detras J, Abriol-Santos JM, Chebotarov D, et al. Rice SNP-seek database update: new SNPs, indels, and queries. Nucleic Acids Res. 2017;45(D1):D1075–81.
26. The 3000 rice genomes project. The 3,000 rice genomes project. GigaScience. 2014;3(1):7.
27. An S, Park S, Jeong D-H, Lee D-Y, Kang H-G, Yu J-H, et al. Generation and analysis of end sequence database for T-DNA tagging lines in rice. Plant Physiol. 2003;133(4):2040–7.
28. Jeon J-S, Lee S, Jung K-H, Jun S-H, Jeong D-H, Lee J, et al. T-DNA insertional mutagenesis for functional genomics in rice. Plant J. 2000;22(6):561–70.
29. Jeong D-H, An S, Park S, Kang H-G, Park G-G, Kim S-R, et al. Generation of a flanking sequence-tag database for activation-tagging lines in *japonica* rice. Plant J. 2006;45(1):123–32.
30. Heidel AJ, Clarke JD, Antonovics J, Dong X. Fitness costs of mutations affecting the systemic acquired resistance pathway in *Arabidopsis thaliana*. Genetics. 2004;168(4):2197–206.
31. Ouibrahim L, Mazier M, Estevan J, Pagny G, Decroocq V, Desbiez C, et al. Cloning of the *Arabidopsis* *rwm1* gene for resistance to watermelon mosaic virus points to a new function for natural virus resistance genes. Plant J. 2014;79:705–16.
32. Poque S, Pagny G, Ouibrahim L, Chague A, Eyraud JP, Caballero M, et al. Allelic variation at the *rpv1* locus controls partial resistance to plum pox virus infection in *Arabidopsis thaliana*. BMC Plant Biol. 2015;15:159.
33. Hebrard E, Pinel-Galzi A, Oludare A, Poulicard N, Aribi J, Fabre S, et al. Identification of a hypervirulent pathotype of rice yellow mottle virus: a threat to genetic resistance deployment in West-Central Africa. Phytopathology. 2018;108(2):299–307.
34. Agnoul Y, Yelome I, Sie M, Albar L, Ghesquière A, Silue D. Resistance of selected *Oryza glaberrima* landraces and their intra-specific breeding lines to Beninese rice yellow mottle virus isolates. Crop Prot. 2019;119:172–6.
35. Thottappilly GG, Rossel HW. Evaluation of resistance to rice yellow mottle virus in *Oryza* species. Indian J Virol. 1993;9(1):65–73.
36. McDowell JM, Simon SA. Recent insights into R gene evolution. Mol Plant Pathol. 2006;7(5):437–48.
37. Guo YL, Fitz J, Schneeberger K, Ossowski S, Cao J, Weigel D. Genome-wide comparison of nucleotide-binding site-leucine-rich repeat-encoding genes in *Arabidopsis*. Plant Physiol. 2011;157(2):757–69.
38. Zhang Y, Guo M, Shen J, Song X, Dong S, Wen Y, et al. Comparative genomics analysis in grass species reveals two distinct evolutionary strategies adopted by R genes. Sci Rep. 2019;9(1):1–10.
39. Karasov TL, Horton MW, Bergelson J. Genomic variability as a driver of plant-pathogen coevolution? Curr Opin Plant Biol. 2014;18:24.
40. Mauricio R, Stahl EA, Korves T, Tian D, Kreitman M, Bergelson J. Natural selection for polymorphism in the disease resistance gene *Rps2* of *Arabidopsis thaliana*. Genetics. 2003;163(2):735–46.
41. Tian D, Araki H, Stahl E, Bergelson J, Kreitman M. Signature of balancing selection in *Arabidopsis*. Proc Natl Acad Sci U S A. 2002;99(17):11525–30.
42. Charron C, Nicolai M, Gallois J-L, Robaglia C, Moury B, Palloix A, et al. Natural variation and functional analyses provide evidence for co-evolution between plant eIF4E and potyviral VPg. Plant J. 2008;54(1):56–68.
43. Pinel-Galzi A, Traoré O, Séré Y, Hébrard E, Fargette D. The biogeography of viral emergence: rice yellow mottle virus as a case study. Curr Opin Virol. 2015;10:7–13.
44. Trovão NS, Baele G, Vrancken B, Bielejec F, Suchard MA, Fargette D, et al. Host ecology determines the dispersal patterns of a plant virus. Virus Evol. 2015;1(1):vev016.
45. Choi JY, Zaidem M, Gutaker R, Dorph K, Singh RK, Purugganan MD. The complex geography of domestication of the African rice *Oryza glaberrima*. PLoS Genet. 2019;15:1–34.
46. Portères R. Berceaux agricoles primaires sur le continent africain. J Afr Hist. 1962;3(2):195–210.
47. Ndjondjop M-N, Semagn K, Gouda AC, Kpeki SB, Dro Tia D, Sow M, et al. Genetic variation and population structure of *Oryza glaberrima* and development of a mini-core collection using DArTseq. Front Plant Sci. 2017; 8:1748.
48. Bouet A, Amancho AN, Kouassi N, Anguete K. Comportement de nouvelles lignées isogéniques de riz irrigué dotées du gène de résistance (*rymv1*) au RYMV en Afrique de l'ouest: situation en Côte d'Ivoire. Int J Biol Chem Sci. 2013;7(3):1221.
49. Ndjondjop MN, Albar L, Sow M, Yao N, Djedatin G, Thiémélé D, et al. Integration of molecular markers in rice improvement: a case study on resistant rice yellow mottle virus. In: Wopereis M, Johnson D, Ahmadi N, Tollens E, Jalloh A, editors. Realizing Africa's rice promise. Wallingford: CAB; 2013. p. 161–72.
50. Pinel-Galzi A, Hébrard E, Traoré O, Silué D, Albar L. Protocol for RYMV inoculation and resistance evaluation in rice seedlings. Bio-protocol. 2018; 8(11):1–13.
51. Kawahara Y, de la Bastide M, Hamilton JP, Kanamori H, McCombie WR, Ouyang S, et al. Improvement of the *Oryza sativa* Nipponbare reference genome using next generation sequence and optical map data. Rice. 2013; 6(1):4.



52. Gigwa Release v2.2. <https://gigwa.ird.fr/gigwa/>. Accessed 9 Sept 2019.
53. Sempéré G, Pétel A, Rouard M, Frouin J, Hueber Y, De Bellis F, Larmande P. Gigwa v2—Extended and improved genotype investigator. *GigaScience*. 2019;8(5):giz051.
54. McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytsky A, et al. The genome analysis toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res*. 2010;20:254–60.
55. den Dunnen JT, Dalgleish R, Maglott DR, Hart RK, Greenblatt MS, McGowan-Jordan J, et al. HGVS recommendations for the description of sequence variants: 2016 update. *Hum Mutat*. 2016;37(6):564–9.
56. Rozen S, Skaletsky H. Primer3 on the WWW for general users and for biologist programmers. In: Misener S, Krawetz SA, editors. *Bioinformatics Methods and Protocols*. Totowa: Humana Press; 1999. p. 365–86.
57. Neff MM, Turk E, Kalishman M. Web-based primer design for single nucleotide polymorphism analysis. *Trends Genet*. 2002;18(12):613–5.

## Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

**Ready to submit your research? Choose BMC and benefit from:**

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

**At BMC, research is always in progress.**

Learn more [biomedcentral.com/submissions](https://biomedcentral.com/submissions)

