

Metatranscriptomics and small RNA analysis revealed that viral covert coinfection resulted in disease symptoms reminiscent of rice sterility

Shufen Chao

Chinese Academy of Medical Sciences & Peking Union Medical College Fuwai Hospital

Haoran Wang

China National Rice Research Institute, Chinese Academy of Agricultural Sciences

Qing Yan

China National Rice Research Institute, Chinese Academy of Agricultural Sciences

Long Chen

Chengdu Branch National Science Library Chinese Academy of Sciences

Guoqing Chen

China National Rice Research Institute, Chinese Academy of Agricultural Sciences

Lijuan Wu

China National Rice Research Institute, Chinese Academy of Agricultural Sciences

Yihong Wu

China National Rice Research Institute, Chinese Academy of Agricultural Sciences

Xudong Zhu

China National Rice Research Institute, Chinese Academy of Agricultural Sciences

Baozhong Meng

University of Guelph

Guozhong Feng (✉ fengguozhong@caas.cn)

China National Rice Research Institute <https://orcid.org/0000-0002-9489-9322>

Research

Keywords: virome, metatranscriptomics, covert infection, virus disease, RNA virus, disease symptoms, small interfering RNA, viral defense

Posted Date: May 1st, 2020

DOI: <https://doi.org/10.21203/rs.3.rs-25471/v1>

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

[Read Full License](#)

Abstract

Background

Viral pathogens are a major threat to stable crop production. The discovery of viral diseases traditionally concerns apparent infection that shows obvious symptoms in crop plants. However, little is known about the covert infection of crop plants by viruses. In this study, we used deep metatranscriptomic sequencing and small RNA analysis to identify covert infection of rice plants by viruses.

Results

Our results showed that introgression of the dominant brown planthopper (BPH) resistance gene *Bph3* into the high-yielding but BPH-susceptible *indica* variety Ms55 via a backcross strategy significantly enhanced resistance to BPH. However, *Bph3*-carrying backcross lines infested by BPH exhibited panicle enclosure and failed to produce seed at the mature stage, which are typical characteristics of sterile rice plants. Using a metatranscriptomic analysis, we identified six RNA viruses in backcross line Rby1 and eleven RNA viruses in backcross line Rby2, including eight novel viruses that fell within existing families and orders. Furthermore, our small RNA analysis revealed the biogenesis of viral small interfering RNAs that represented active virus infection in rice plants.

Conclusion

Bph3-carrying backcross rice lines are resistant to BPH but are susceptible to viral infection. We identified viral covert coinfection in sterile rice plants by deep metatranscriptomic sequencing and small RNA analysis. Our results suggested that covert coinfection of rice plants by RNA viruses resulted in disease symptoms reminiscent of rice sterility. To develop rice varieties resistant to BPH, it is necessary to introgress genes resistant to not only BPH but also viral infection.

Background

Rice is one of the most important cereal crop species and sustains half of the global population. Rice is cultivated in various environments where a hot and humid climate favors not only rice growth but also virus propagation and spread. The first major outbreak of rice viruses was recorded for rice dwarf virus (RDV) in Japan [1]. To date, at least 22 viruses have been reported in rice plants, including those from the *Reoviridae*, *Benyviridae*, *Rhabdoviridae* and *Phenuiviridae* families [1-5]. These viruses result in rice diseases that pose serious threats to stable rice production. Rice plants infected by viruses showed symptoms such as pronounced stunting, dark green leaves, increased tillering and failure to elongate [1]. The majority of rice viruses are transmitted by rice-infesting insects, including planthoppers and leafhoppers [1, 4, 6-8] {Ding, 2007 #38}.

The discovery of viruses traditionally concerns those that are pathogenic to their hosts and that can be isolated; these viral infections cause severe diseases and show obvious symptoms [1, 5, 9, 10]. However,

many virus infections are asymptomatic or induce unclear symptoms in their animal or plant hosts; these viruses may accumulate to relatively low titers in their host organisms or become latent such that virus production ceases [11-17]. Consequently, they cannot easily be isolated or cultured from their hosts using traditional methods. Metatranscriptomics has proven to be a powerful approach to uncover hidden viruses in many organisms, including humans, arthropods, and plants [13, 14, 18-21].

Metatranscriptomics provides sufficient coverage to reconstruct complete viral genomes and allows a straightforward characterization of viral diversity [22]{Hibino, 1996 #16}. However, a metatranscriptomic survey of viral diversity has not been performed on cereal crop species, including rice.

The brown planthopper (BPH, *Nilaparvata lugens* Stål) is the most destructive pest of rice (*Oryza sativa*) and is a substantial threat to rice production, causing losses of billions of dollars annually [23, 24]. The development of resistant rice cultivars is considered the most economically effective and environmentally friendly method of controlling this insect [25]. Ms55 is a high-yielding but BPH-susceptible cultured elite *indica* cultivars in China. To develop BPH-resistant *indica* cultivars, the BPH resistance gene *Bph3* from cultivar TZ21 was introgressed into Ms55 to produce an initial backcross population (BC₁F) [23]. To test whether the BC₁F rice plants are resistant to BPH, *Bph3*-carrying Rby1 and Rby2 were selected from BC₁F and grown in a greenhouse. Our results revealed that both rice lines infested by BPH could grow as healthy rice plants, while they were sterile at the mature stage. They also did not show any symptoms of fungal or bacterial infections. We tried to determine whether viral infection caused rice sterility by deep metatranscriptomic sequencing. Unexpectedly, we discovered thirteen viruses in both rice lines, including eight novel viruses and five known viruses, and small RNA analysis revealed active viral infection in the rice plants. Taken together, our results suggested that covert coinfection of rice plants by RNA viruses resulted in disease symptoms reminiscent of rice sterility.

Results

***Bph3*-carrying backcross rice lines infested by BPH are sterile**

Ms55 is a high-yielding elite *indica* cultivar in China, but it is susceptible to BPH. To develop BPH-resistant *indica* cultivars, Ms55 was used as the recurrent parent and backcrossed with TZ21 harboring the resistance gene *Bph3* to produce an initial backcross (BC₁F) population (Fig. S1). Two *Bph3*-carrying *indica* lines, Rby1 and Rby2, were selected from the BC₁F population.

To test whether the introgression of *Bph3* can improve resistance to BPH, rice plants from the Rby1 and Rby2 lines grew under greenhouse conditions, and each line at the seedling stage was infested with BPH of mixed biotypes (with biotype 2 being the dominant one; the BPH biotypes refer to specific populations of BPH classified according to their virulence on different BPH resistance genes) [24] collected from rice fields in Hangzhou. Ms55 was used as negative control, and TZ21 was regarded as positive control. The *Bph3*-carrying Rby1, Rby2 and TZ21 lines were not visibly damaged, while BPH infestation resulted in 100% mortality in the susceptible recurrent parent Ms55 at 19 days. When Rby1, Rby2, TZ21 and Ms55 were infested with a fixed number of BPHs, the BPH population on Ms55 increased steadily over time. In

contrast, the BPH population on Rby1, Rby2, and TZ21 decreased dramatically within a few hours of infestation. These observations indicated that Rby1 and Rby2 displayed strong resistance to BPH. Although the behavior of Rby1 and Rby2 infested by BPH was similarly to that of Ms55 or TZ21 in terms of their vegetative growth, both rice lines showed panicle enclosure, and only a few seeds were produced on each rice plant, which are typical characteristics of sterile rice plants (Fig. 1c-e); however, the agronomic traits of the Rby1 or Rby2 rice plants not infested by BPH were similar to those of Ms55 or TZ21, including earing and seed formation (Fig. 1a-b).

Multiple RNA viruses were identified in sterile rice plants by deep metatranscriptomic sequencing

Rice plants of lines Rby1 or Rby2 not infested by BPH were as healthy as those of Ms55 or TZ21 (Fig. 1a-b). However, those of lines Rby1 and Rby2 became sterile when they were infested by BPH, and these sterile rice plants did not exhibit any disease symptoms indicative of fungal or bacterial infections or typical characteristics of viral infection, such as pronounced stunting and dark green leaves. The only difference between the healthy and sterile rice plants was BPH infestation; it is known that BPH can transmit viruses as vectors to rice plants, but we did not isolate any viruses from the sterile rice plants. We speculated whether viral covert infection caused rice sterility. To test this hypothesis, we performed deep metatranscriptomic sequencing of rice plants, including Rby1-21 from line Rby1 and Rby2-45 from line Rby2. RNA sequencing of rRNA-depleted libraries yielded 12.69 Gb of data for Rby1-21 and 10.64 Gb of data for Rby2-45, which resulted in 70,923,042 and 84,627,146 reads, respectively. We de novo assembled these reads into 328,146 contigs for Rby1-21 and 586,111 contigs for Rby2-45 (Table 1).

We first examined the assembled contigs that matched previously characterized RNA viruses. We identified five known viruses in Rby1-21 and Rby2-45. Rice tombus-like virus 1 (RTV1) and rice ragged stunt virus (RRSV) were present in both Rby1-21 and Rby2-45, while rice picorna-like virus 1 (RPIV1), rice toti-like virus (RToV) and a brown planthopper virus, *Nilaparvata lugens* reovirus (NLRV), were present only in Rby2-45 (Table 1).

Next, we identified virus-like sequences by BLASTing against reference sequences from viral genomes available in GenBank. We discovered eight novel viruses in both rice lines (Fig. 2), including three negative-sense RNA viruses and five positive-sense RNA viruses (Table S1). Of the eight RNA viruses, rice mononega-like virus (RMV) and rice peribunya-like virus (RPeV) were present in both Rby1-21 and Rby2-45 (Table S1). Rice picorna-like virus 2 (RPIV2) and rice noda-like virus (RNV) were present only in Rby1-21 (Table 1 and Table S1). Rice phasma-like virus (RPhV), rice tombus-like virus 2 (RTV2), rice tombus-like virus 3 (RTV3) and Rice picorna-like virus 3 (RPIV3) were present only in Rby2-45 (Table 1 and Table S1). Thus, we identified thirteen RNA viruses from Rby1-21 and Rby2-45 in total, including five known RNA viruses and eight novel RNA viruses (Table 1). We detected six RNA viruses in Rby1-21 and eleven RNA viruses in Rby2-45 (Table 1). Of these viruses, four RNA viruses were present in both Rby1-21 and Rby2-45, two RNA viruses were present in only Rby1-21, and seven RNA viruses were present in only Rby2-45 (Fig. S2). However, we did not detect any viruses in Rby1-N65 from line Rby1 or in Rby2-N32 from line Rby2, which were not infested by BPH.

We identified three novel negative-sense RNA viruses (Table S1). Of these viruses, RPeV is closely related to *Penicillium roseopurpureum* negative ssRNA virus 1, RPhV is related to *Anopheles triannulatus* orthophasmavirus, and RMV is similar to Tacheng tick virus 5 (Table S1). In the RdRp phylogeny, RPeV was clustered within the family *Peribunyaviridae*, and RPhV belonged to the family *Phasmaviridae*, both of the order *Bunyvirales* (Fig. 3). RMV clustered within a currently unclassified family of the order *Mononegavirales* (Fig. 3 and Fig. S3).

We also discovered five undescribed positive-sense RNA viruses in this study (Table S1). In the RdRp phylogeny, these viruses fell within the *Tombusviridae*, *Picornaviridae* and *Nodaviridae* families (Fig. 3). Two positive-sense RNA viruses, RTV2 and RTV3, belonged to the family *Tombusviridae* (Fig. 3). RTV2 was closely related to soybean leaf-associated ssRNA virus 1, and RTV3 was similar to *Setosphaeria turcica* ambiguvirus 1 (Table S1). Two positive-sense RNA viruses, RPiV2 and RPiV3, fell within the family *Picornaviridae* (Fig. 3); RPiV2 was closely related to *Picornavirales* sp., and RPiV3 was related to Hubei picorna-like virus 35 (Table S1). The remaining RNV clustered within the family *Nodaviridae* (Fig. 3), and RNV was closely related to Hubei orthoptera virus 4 (Table S1).

Although phylogenetic analysis showed that these sequences are viral in origin, they may represent endogenous viral elements (EVEs) integrated into the rice plant genome [11] or may be derived from surface contamination rather than active infection. To exclude the possibility that these contigs represent EVEs segregating in rice plants, we mapped the raw reads from the Rby1-21 and Rby2-45 genomes to our set of candidate viruses revealed by BLAST and confirmed that no genome mapped at a rate high enough to be consistent with a genome copy of any virus in a particular individual.

Small RNA analysis indicated active virus infections in sterile rice plants

Rice plants utilize small RNA pathways for viral defense [26, 27]. If a virus is active in rice plants, this is often indicated by the presence of an antiviral immune response [28]. Thus, we determined the presence of virus-derived small interfering RNAs (vsiRNAs) in rice plants as being indicative of active viral infection. We constructed small RNA libraries from Rby1-21 and Rby2-45 samples. Both libraries were subjected to 50 bp single-end sequencing, resulting in 24,822,021 and 23,341,707 reads for Rby1-21 and Rby2-45, respectively.

To map the resulting small RNA reads to the putative viruses described above, we first removed the small RNA reads that are specific to rice plants and then aligned the reads to the viral genomes described in Table 1. The results showed that differentially abundant small RNAs were mapped to putative viruses (Table S2), and an abundance of small RNAs were identified for **RTV1 (134,715 reads) and RRSV (7,108 reads) from Rby1-21 as well as RTV1 (109,437 reads), RRSV (23,558 reads) and RPeV (3,639 reads) from Rby2-45**. Small RNAs with fewer than 200 reads were excluded because of possible random degradation.

To test whether these small RNA sequences were derived from active virus infection, we analyzed all 18-30 nt small RNAs from identified viruses in Rby1-21 and Rby2-45. VsiRNAs detected in plants infected with RNA viruses are predominantly 21 nucleotides in length produced by Dicer-like 4 (DCL4), and 22-

nucleotide vsRNAs are produced by DCL 2 [29, 30]. Our small RNA reads have a size distribution of 21 to 22 nt for RTV1 and RRSV in Rby1-21 as well as RPeV, RTV1, and RRSV in Rby2-45. These small RNAs occur mainly in the sense and antisense orientations (Fig. 4), indicating that the vsRNAs were produced from double-stranded RNA replicative intermediates and that Dicer-like enzymes cleaved double-stranded RNAs into vsRNA. The cleaved double-stranded vsRNAs are bound and sorted by AGOs depending on the 5'-terminal base [31]. To assess the base preference at the 5' terminus in the rice samples, we calculated the base percentage for the vsRNAs. As shown for the three viruses (RTV1, RRSV and RPeV) in Fig. 4, base G was the least favored at the 5'-terminal nucleotide for all viruses, and the viruses showed a distinct preference for the remaining three bases at the 5' terminus. Base U was enriched in 21-nt vsRNAs for RPeV and RRSV, while base C was enriched in those for RTV1. These results were consistent with those of a previous study [32]. Twenty-one-nucleotide vsRNAs occurred in regions of identified genomic RNAs (Fig. 5), and those derived from antisense strands showed a high proportion for each virus (Table S3). Collectively, our data demonstrated that RTV1 and RRSV in Rby1-21 as well as RPeV, RTV1 and RRSV in Rby2-45 were active infections.

Discussion

Viruses are a major threat to global rice production. In this report, we described a diverse set of new viruses in rice lines Rby1 and Rby2 through deep transcriptomic sequencing. We identified and assembled thirteen complete or nearly complete genomic sequences, eight of which were previously undescribed, including three negative-sense RNA viruses and five positive-sense RNA viruses. We also identified five known viruses containing two positive-strand viruses, RTV1 and RPiV1, as well as three double-stranded viruses, RRSV, NLRV and RTov [1, 33, 34]. Notably, RPiV1 and RTov were first discovered in the white-backed planthopper *Sogatella furcifera* [34, 35]. In this report, both viruses were identified in Rby2-45, indicating that they have more hosts and that BPH might harbor RPiV1 and RTov and transmit them to Rby2-45. Metatranscriptomic approaches have been successfully implemented to discover highly diverse and novel viruses from mosquito, bat, and *Drosophila* samples [20, 22, 36, 37]. To the best of our knowledge, this is the first comprehensive high-throughput survey of viral sequences associated with rice plants.

Breeding for resistant rice cultivars is one of the most important approaches to control insect pests [38-40]. In this report, we used Ms55, an *indica* variety susceptible to BPH, to backcross with TZ21 harboring the resistance gene *Bph3* to produce a BC₁F population. During a 5-year period (2013-2017), the *Bph3*-carrying rice lines from the BC₁F population were as healthy as TZ21 in 2013, 2015 and 2016, while the same rice lines were sterile in 2014 and 2017. We carefully sought to explain the rice sterility that had been occurring since 2014. In our study, these sterile rice plants were infested by BPH, an important vector that can harbor and transmit viruses to rice plants, and we speculated that virus infection may cause rice sterility. However, we have not isolated any viruses from these sterile rice lines, indicating that viral pathogen(s) could not be identified by traditional approaches. We then tried to use deep metatranscriptomic sequencing to search for possible viral pathogen(s), and we in turn identified thirteen

RNA viruses. Small RNA analysis revealed that RTV1 and RRSV in Rby1-21 as well as RPeV, RTV1 and RRSV in Rby2-45 were active infections. Although Rby1 and Rby2 are resistant to BPH, they are susceptible to viral infection. Thus, to develop rice varieties resistant to BPH, it is necessary to introgress genes resistant to not only BPH but also viral infection.

To confirm that covert coinfection of rice plants by RNA viruses is associated with sterile symptoms, we showed several lines of evidence: (i) Both rice lines from the BC₁F population were sterile only when they were infested by BPH, while rice plants not infested by BPH grew normally, as did the healthy TZ21 plants, and it is known that BPH can harbor viruses and transmit them to rice plants; (ii) We identified multiple viruses in each line by deep metatranscriptomic sequencing, and vsRNA analysis revealed that RTV1 and RRSV in Rby1-21 as well as RTV1, RRSV and RPeV in Rby2-45 were active infections; (iii) We searched for contigs that would match the DNA polymerase of DNA viruses, but we did not find any sequence related to viral DNA polymerases; and (iv) We also reasoned that contigs that lack similarity searches to reference sequences but that display a signature of DCL2, DCL3 or DCL4 processing (high levels of 21-24 nt vsRNAs) may also be viral in origin [41, 42]. Using these small RNA criteria, we did not identify any contig that is potentially viral in origin. Additionally, the agronomic traits of the backcross lines not infested by BPH were similar to those of Ms55 and TZ21, including earing and seed formation, indicating that these rice plants were not infected by pathogens, including bacterial or fungal ones. Healthy and sterile rice plants from lines Rby1 and Rby2 grew under the same greenhouse conditions, suggesting that the possibility that fungal or bacterial infection caused rice sterility was excluded. Thus, rice sterility was derived from coinfections of RTV1 and RRSV in Rby1-21 as well as RTV1, RRSV and RPeV in Rby2-45. Taken together, our results suggest that covert coinfection of rice plants by RNA viruses rather than rice hybridization resulted in rice sterility.

In this report, we did not isolate any virus from infected rice plants despite of extensive efforts. A virus that exhibited high (85%) nucleotide sequence similarity to RTV1, *Dianthovirus* RNA1-like RNA (DR1L RNA), was found in grassy stunt-diseased rice plants together with rice grassy stunt virus (RGSV) [43]. Although the nucleotide sequence of DR1L RNA was identified from cDNA library of grassy stunt-diseased rice leaf tissues, no virus-like particle was observed on fractions from a sucrose density gradient, or no virus was isolated from infected rice plants [43]. In our study, RTV1 identified from both Rby1 and Rby2 showed similar results. RTV1 and DR1L RNA were detected in samples with viral mixed infections, whereas individual RTV1 or DR1L RNA isolate has not been reported to date. RTV1 and DR1L RNA belong to the genus *Dianthovirus* of the family *Tombusviridae*, and members of the genus *Dianthovirus* contain two segments in their genomes. However, RTV1 and DR1L RNA have only RNA1, another RNA2 did not be found in infected samples, indicating that RNA2 was lost. RTV1 showed active infection in both Rby1 and Rby2 even if RNA2 was absent in rice samples, which suggested that RTV1 may be rescued by other viruses in Rby1 and Rby2.

In a mixed infection of two or more type of plant viruses, there is the potential for interaction [44]. If the viruses are in competition with one another, one of the viruses is likely to win and become the predominant virus, whereas if they are cooperating, they are more likely to remain as a mixed infection

[44]. In this report, we demonstrated that RTV1 and RRSV in Rby1-21 as well as RTV1, RRSV and RPeV in Rby2-45 were active, while other viruses in Rby1 and Rby2 showed latent infections, these may be the results of virus-virus interactions. Our results also showed that RPeV in Rby2-45 is active, while that in Rby1-21 is not. A transition from latency to activation occasionally occurs during mixed infections [11, 45]. RPeV may be activated by direct or indirect transactivation of viral gene(s) of latent virus that was absent in Rby1 but presence in Rby2. Thus, further analysis of virus-virus interaction will be necessary to understand the mechanism of virus infection in rice plants.

Conclusion

In this report, two backcross lines, Rby1 and Rby2, infested by BPH, showed typical characteristics of sterile rice plants. We used deep metatranscriptomic sequencing to identify thirteen RNA viruses, including six RNA viruses in Rby1-21 and eleven RNA viruses in Rby2-45, eight of which are newly described viruses. Small RNA analysis showed that RTV1 and RRSV in Rby1-21 as well as RTV1, RRSV and RPeV in Rby2-45 are active infections. Covert coinfection of rice plants by RNA viruses resulted in disease symptoms reminiscent of those of rice sterility. To develop rice varieties resistant to BPH, it is necessary to introgress genes that provide resistance to not only BPH but also viral infection.

Materials And Methods

Plant materials and sample collection

The rice cultivars Ms55 and TZ21 used in this work were obtained from the Chinese Rice Germplasm Resources Center (Hangzhou, China). Ms55, a high-yielding elite *indica* cultivar, was used as a recurrent parent. TZ21, harboring the resistance gene *Bph3*, was used as the donor parent [23]. Two *Bph3*-carrying lines, Rby1 and Rby2, were selected from the BC₁F population by a backcross between the recurrent parent Ms55 and the donor TZ21. Rice plants from lines Rby1, Rby2, Ms55 and TZ21 were grown in a greenhouse at the China National Rice Research Institute, Hangzhou, China. Leaf and stem samples of rice plants from Rby1, Rby2, Ms55 and TZ21 were harvested and placed on dry ice until they were moved to the laboratory, where the samples were subsequently stored at -80°C prior to RNA isolation.

Brown planthopper maintenance

Brown planthoppers were collected from a rice field at the China National Rice Research Institute, Hangzhou (where the BPHs consisted of mixed biotypes and mainly biotype 2), and maintained on the susceptible cultivar Ms55 under greenhouse conditions at the China National Rice Research Institute.

Evaluation of brown planthopper resistance

Bioassays were conducted according to the methods described by Liu, with slight modifications [23]. Plant seeds were germinated in petri dishes, and 30 seeds from each individual plant were sown in a 10 cm-diameter plastic pot with a hole at the bottom. Seven days after sowing, the seedlings were thinned to

20 plants per pot. At the second-leaf stage, the seedlings were infested with second- and third-instar BPH nymphs of 10 insects collected from rice fields per seedling. When all the Ms55 plants had died, the seedling mortality of the other cultivars or lines was recorded. Three replicates were used for each cultivar or line.

Sample processing and sequencing

The cDNA libraries were constructed from Rby1 and Rby2 rice samples. To isolate the total RNA, the samples were homogenized in 700 µl of lysis buffer by TissueRuptor (Qiagen). Total RNA was extracted by using an RNeasy Plus Mini Kit according to the manufacturer's protocol (Qiagen). The quality of the resultant RNA was evaluated using an Agilent 2100 bioanalyzer (Agilent Technologies) before library construction. The host RNA was removed by using a Ribo-Zero-Gold (Human-Mouse-Rat) Kit (Illumina). Two sequencing libraries were constructed by using an Illumina TruSeq Total RNA library preparation kit. Paired-end (150 bp) sequencing of each library was then performed on a HiSeq 2000 platform (Illumina). Library preparation and sequencing were subsequently carried out by Novogene.

Sequence read assembly and virus discovery

Sequencing reads were assembled de novo by using Trinity [46]. The assembled contigs were first compared against the database of all reference RNA virus proteins downloaded from GenBank by using BLASTX, with an E value cutoff at 1E-5 to maximize sensitivity while minimizing false-positive results. The resultant contigs were then compared to information in the nonredundant nucleotide (nt) and protein (nr) databases to remove nonviral sequences. We also performed domain-based BLAST to detect highly divergent viruses. The assembled contigs were compared to the information within the Conserved Domain Database (CDD) version 3.16, with an expected value threshold of 1E-2. The quality-filtered virus contigs with unassembled overlaps were then merged using SeqMan implemented in the Lasergene software package version 7.1 (DNASTar). To confirm the assembly results, the reads were mapped back to the virus genomes with Bowtie2 and then inspected by using Integrated Genomics Viewer (IGV) to detect any assembly errors. The final sequences of the virus genomes were obtained from the majority consensus of the mapping assembly. SAMtools was subsequently used to determine the sequencing depth and coverage [47].

Virus genome annotation

The potential ORFs of the newly identified virus genomes were annotated based on the predicted amino acid sequences and conserved positions in the genome compared to those of the closest related virus genome available in GenBank. Functional domains within each ORF were identified by using BLAST against the CDD, with an expected value threshold of 1E-5.

For viruses with multiple RNA segments, we used various strategies to search for viral genome segments, as described previously [14, 22]. Non-RdRp segments were identified by homology to the proteins of related reference viruses. Other potential segments that displayed no homology to sequences in the

database were identified by using an *in silico* approach that utilizes information on RNA quantity, protein structure, and/or conserved genomic termini. To determine which segments belong to the same virus, we checked (i) the sequence depth of the segments, (ii) the presence of conserved regulatory sequences in the noncoding regions located at termini of the viral genomes, and (iii) the phylogenetic positions of related viral proteins.

We used RT-PCR and Sanger sequencing to verify the presence of each putative virus in the samples. Primers were designed on the basis of the contigs assembled from next-generation sequencing.

Estimation of viral transcript frequency

Viral transcript frequency were estimated according to the methods described previously [14]. To determine the frequency of viral RNAs, we estimated the percentage of reads that mapped to viral RNA within the transcriptome of each sample. To reduce any bias caused by the unequal efficiency of rRNA removal during library preparation, we first removed reads that mapped to rRNA contigs from each library, the remaining reads were then mapped to the entire collection of virus sequences within the library, from which we calculated the overall percentage of viral reads. To confirm that viral RNA comprised a large percentage of transcripts within the sample transcriptome, we re-isolated the total RNA from aliquots of the original homogenates and performed RNA-seq library preparation and sequencing as described above.

Phylogenetic analysis

To determine the phylogenetic relationships of the newly identified RNA viruses, the amino acid sequences of the viral RNA-dependent RNA polymerase (RdRp) proteins identified in this study and those retrieved from GenBank were aligned to infer their evolutionary relationship [14]. The viral RdRp sequences were then aligned using the E-INS-I algorithm in MAFFT (version 7.429) [48]. The sequence alignment was limited to conserved domains, and all ambiguously aligned regions were removed by using TrimAl [49]. For each sequence alignment, the best-fit model of amino acid substitutions was determined by using ProTest version 3.4 [50]. Phylogenetic trees were then constructed by using the maximum likelihood (ML) method implemented in PhyML version 3.0 [51], employing the best-fit substitution model and the Subtree Pruning and Regrafting (SPR) branch-swapping program. The branch support for nodes on the phylogenetic tree was accessed by using an approximate likelihood ratio test (aLRT) with the Shimodaira-Hasegawa-like procedure implemented in PhyML [51].

Small RNA library construction and sequencing

One microgram of total RNA from each rice sample was used to generate a library of small RNAs by using a TruSeq Small RNA Kit (Illumina) according to the sample preparation instructions. Both Rby1-21 and Rby2-45 libraries were used for 50-bp single-end sequencing by the Illumina HiSeq 2000 platform in two lanes.

Small RNA analysis

Bioinformatic analysis of the small RNA data was performed using the CLC Genomic Workbench software package (Qiagen). Briefly, small RNA reads were quality checked, and low-quality reads and adapter sequences were first removed from the raw small RNA data set. Trimmed small RNA sequences shorter than 15 nucleotides were discarded. The remaining reads were mapped to the rice genome to remove host-related reads, and the unmapped reads were subsequently mapped to putative viruses with the same stringency settings.

Declarations

Availability of data and materials

All viral genome sequences generated in this study have been deposited in the GenBank database under accession numbers MT317153-MT317177. The raw sequence data reads are available at the NCBI Sequence Read Archive (SRA) database under BioProject accession number (PRJNA612173).

Acknowledgments

We thank P. Krell for the scientific discussions and Q. Bai and H. Wang for their excellent technical support.

Funding

This research was supported by the National Natural Science Foundation of China (Nos. 31572006, 31701847 and 31972983), the Innovation Project of the Chinese Academy of Agricultural Sciences and the Central Level Public Interest Research Institute for Basic R & D Special Fund Business (No. 2017RG002-7).

Author information

Affiliations

State Key Laboratory of Rice Biology, China National Rice Research Institute, Chinese Academy of Agricultural Sciences, Hangzhou, China

Shufen Chao, Haoran Wang, Qing Yan, Long Chen, Guoqing Chen, Lijuan Wu, Yihong Wu, Xudong Zhu, Guozhong Feng

Department of Molecular and Cellular Biology, University of Guelph, Guelph, ON, Canada.

Baozhong Meng

Contributions

GF and SC conceived the ideas, and GF, BM and SC designed the experiments. SC, HW, QY, GC, LC, LW, YW and GF analyzed the data. SC, HW, QY, GC performed the experiments. XZ and LC performed the

experiments with the rice hybrids. GF, BM and SC wrote the manuscript.

Corresponding author

Correspondence to Guozhong Feng

Ethics declarations

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing financial interests.

References

1. Hibino H: **Biology and epidemiology of rice viruses.** *Annu Rev Phytopathol* 1996, **34**:249-274.
2. Zhou G, Xu D, Xu D, Zhang M: **Southern rice black-streaked dwarf virus: a white-backed planthopper-transmitted fijivirus threatening rice production in Asia.** *Front Microbiol* 2013, **4**:270.
3. Wei T, Li Y: **Rice Reoviruses in Insect Vectors.** *Annu Rev Phytopathol* 2016, **54**:99-120.
4. Liu W, Hajano JU, Wang X: **New insights on the transmission mechanism of tenuiviruses by their vector insects.** *Curr Opin Virol* 2018, **33**:13-17.
5. Chen S, Li W, Huang X, Chen B, Zhang T, Zhou G: **Symptoms and yield loss caused by rice stripe mosaic virus.** *Virol J* 2019, **16**(1):145.
6. Otuka A: **Migration of rice planthoppers and their vectored re-emerging and novel rice viruses in East Asia.** *Front Microbiol* 2013, **4**:309.
7. Jia DS, Chen Q, Mao QZ, Zhang XF, Wu W, Chen HY, Yu XZ, Wang ZQ, Wei TY: **Vector mediated transmission of persistently transmitted plant viruses.** *Curr Opin Virol* 2018, **28**:127-132.
8. Chen Q, Wei TY: **Viral receptors of the gut: insect-borne propagative plant viruses of agricultural importance.** *Curr Opin Insect Sci* 2016, **16**:9-13.
9. Zhou GH, Wen JJ, Cai DJ, Li P, Xu DL, Zhang SG: **Southern rice black-streaked dwarf virus: A new proposed Fijivirus species in the family Reoviridae.** *Chinese Sci Bull* 2008, **53**(23):3677-3685.
10. Yang X, Huang J, Liu C, Chen B, Zhang T, Zhou G: **Rice Stripe Mosaic Virus, a Novel Cytorhabdovirus Infecting Rice via Leafhopper Transmission.** *Front Microbiol* 2016, **7**:2140.

11. Takahashi H, Fukuhara T, Kitazawa H, Kormelink R: **Virus Latency and the Impact on Plants.** *Front Microbiol* 2019, **10**:2764.
12. Barba M, Czosnek H, Hadidi A: **Historical perspective, development and applications of next-generation sequencing in plant virology.** *Viruses* 2014, **6**(1):106-136.
13. Li CX, Shi M, Tian JH, Lin XD, Kang YJ, Chen LJ, Qin XC, Xu J, Holmes EC, Zhang YZ: **Unprecedented genomic diversity of RNA viruses in arthropods reveals the ancestry of negative-sense RNA viruses.** *Elife* 2015, **4**.
14. Shi M, Lin XD, Tian JH, Chen LJ, Chen X, Li CX, Qin XC, Li J, Cao JP, Eden JS *et al*: **Redefining the invertebrate RNA virosphere.** *Nature* 2016, **540**(7634):539-543.
15. Kamitani M, Nagano AJ, Honjo MN, Kudoh H: **RNA-Seq reveals virus-virus and virus-plant interactions in nature.** *FEMS Microbiol Ecol* 2016, **92**(11).
16. Stobbe AH, Roossinck MJ: **Plant virus metagenomics: what we know and why we need to know more.** *Front Plant Sci* 2014, **5**:150.
17. Roossinck MJ, Martin DP, Roumagnac P: **Plant Virus Metagenomics: Advances in Virus Discovery.** *Phytopathology* 2015, **105**(6):716-727.
18. Ng TF, Kondov NO, Deng X, Van Eenennaam A, Neiberghs HL, Delwart E: **A metagenomics and case-control study to identify viruses associated with bovine respiratory disease.** *J Virol* 2015, **89**(10):5340-5349.
19. Tokarz R, Williams SH, Sameroff S, Sanchez Leon M, Jain K, Lipkin WI: **Virome analysis of Amblyomma americanum, Dermacentor variabilis, and Ixodes scapularis ticks reveals novel highly divergent vertebrate and invertebrate viruses.** *J Virol* 2014, **88**(19):11480-11492.
20. Webster CL, Waldron FM, Robertson S, Crowson D, Ferrari G, Quintana JF, Brouqui JM, Bayne EH, Longdon B, Buck AH *et al*: **The Discovery, Distribution, and Evolution of Viruses Associated with Drosophila melanogaster.** *PLoS Biol* 2015, **13**(7):e1002210.
21. Zhang YZ, Shi M, Holmes EC: **Using Metagenomics to Characterize an Expanding Virosphere.** *Cell* 2018, **172**(6):1168-1172.
22. Shi M, Neville P, Nicholson J, Eden JS, Imrie A, Holmes EC: **High-Resolution Metatranscriptomics Reveals the Ecological Dynamics of Mosquito-Associated RNA Viruses in Western Australia.** *J Virol* 2017, **91**(17).
23. Liu Y, Wu H, Chen H, Liu Y, He J, Kang H, Sun Z, Pan G, Wang Q, Hu J *et al*: **A gene cluster encoding lectin receptor kinases confers broad-spectrum and durable insect resistance in rice.** *Nat Biotechnol* 2015, **33**(3):301-305.
24. Cheng XY, Zhu LL, He GC: **Towards Understanding of Molecular Interactions between Rice and the Brown Planthopper.** *Mol Plant* 2013, **6**(3):621-634.
25. Khush GS: **Green revolution: the way forward.** *Nat Rev Genet* 2001, **2**(10):815-822.
26. Guo Z, Li Y, Ding SW: **Small RNA-based antimicrobial immunity.** *Nat Rev Immunol* 2019, **19**(1):31-44.
27. Yang Z, Li Y: **Dissection of RNAi-based antiviral immunity in plants.** *Curr Opin Virol* 2018, **32**:88-99.

28. Ding SW: **RNA-based antiviral immunity.** *Nat Rev Immunol* 2010, **10**(9):632-644.
29. Deleris A, Gallego-Bartolome J, Bao JS, Kasschau KD, Carrington JC, Voinnet O: **Hierarchical action and inhibition of plant Dicer-like proteins in antiviral defense.** *Science* 2006, **313**(5783):68-71.
30. Liu B, Chen Z, Song X, Liu C, Cui X, Zhao X, Fang J, Xu W, Zhang H, Wang X *et al.*: **Oryza sativa dicer-like4 reveals a key role for small interfering RNA silencing in plant development.** *Plant Cell* 2007, **19**(9):2705-2718.
31. Mi SJ, Cai T, Hu YG, Chen Y, Hodges E, Ni FR, Wu L, Li S, Zhou H, Long CZ *et al.*: **Sorting of small RNAs into Arabidopsis argonaute complexes is directed by the 5' terminal nucleotide.** *Cell* 2008, **133**(1):116-127.
32. Xu CX, Sun XP, Taylor A, Jiao C, Xu YM, Cai XF, Wang XL, Ge CH, Pan GH, Wang QX *et al.*: **Diversity, Distribution, and Evolution of Tomato Viruses in China Uncovered by Small RNA Sequencing.** *Journal of Virology* 2017, **91**(11).
33. Miranda GJ, Aliyari R, Shirako Y: **Nucleotide sequence of a Dianthovirus RNA1-like RNA found in grassy stunt-diseased rice plants.** *Arch Virol* 2001, **146**(2):225-238.
34. Wu N, Zhang P, Liu W, Cao M, Wang X: **Sequence analysis and genomic organization of a new insect iflavirus, Sogatella furcifera honeydew virus 1.** *Arch Virol* 2018, **163**(7):2001-2003.
35. Zhang P, Liu W, Cao M, Massart S, Wang X: **Two novel totiviruses in the white-backed planthopper, Sogatella furcifera.** *J Gen Virol* 2018, **99**(5):710-716.
36. Hayman DTS: **Bats as Viral Reservoirs.** *Annu Rev Virol* 2016, **3**:77-99.
37. Wu Z, Yang L, Ren X, He G, Zhang J, Yang J, Qian Z, Dong J, Sun L, Zhu Y *et al.*: **Deciphering the bat virome catalog to better understand the ecological diversity of bat viruses and the bat origin of emerging infectious diseases.** *ISME J* 2016, **10**(3):609-620.
38. Jena KK, Jeung JU, Lee JH, Choi HC, Brar DS: **High-resolution mapping of a new brown planthopper (BPH) resistance gene, Bph18(t), and marker-assisted selection for BPH resistance in rice (Oryza sativa L.).** *Theor Appl Genet* 2006, **112**(2):288-297.
39. Liu Y, Chen L, Liu Y, Dai H, He J, Kang H, Pan G, Huang J, Qiu Z, Wang Q *et al.*: **Marker assisted pyramiding of two brown planthopper resistance genes, Bph3 and Bph27 (t), into elite rice Cultivars.** *Rice (N Y)* 2016, **9**(1):27.
40. Jiang H, Hu J, Li Z, Liu J, Gao G, Zhang Q, Xiao J, He Y: **Evaluation and breeding application of six brown planthopper resistance genes in rice maintainer line Jin 23B.** *Rice (N Y)* 2018, **11**(1):22.
41. Aguiar ER, Olmo RP, Paro S, Ferreira FV, de Faria IJ, Todjro YM, Lobo FP, Kroon EG, Meignin C, Gatherer D *et al.*: **Sequence-independent characterization of viruses based on the pattern of viral small RNAs produced by the host.** *Nucleic Acids Res* 2015, **43**(13):6191-6206.
42. Wu Q, Luo Y, Lu R, Lau N, Lai EC, Li WX, Ding SW: **Virus discovery by deep sequencing and assembly of virus-derived small silencing RNAs.** *Proc Natl Acad Sci U S A* 2010, **107**(4):1606-1611.
43. Miranda GJ, Aliyari R, Shirako Y: **Nucleotide sequence of a Dianthovirus RNA1-like RNA found in grassy stunt-diseased rice plants.** *Arch Virol* 2001, **146**(2):225-238.

44. Roossinck MJ: **Symbiosis versus competition in plant virus evolution.** *Nat Rev Microbiol* 2005, **3**(12):917-924.
45. DaPalma T, Doonan BP, Trager NM, Kasman LM: **A systematic approach to virus-virus interactions.** *Virus Res* 2010, **149**(1):1-9.
46. Grabherr MG, Haas BJ, Yassour M, Levin JZ, Thompson DA, Amit I, Adiconis X, Fan L, Raychowdhury R, Zeng QD *et al*: **Full-length transcriptome assembly from RNA-Seq data without a reference genome.** *Nature Biotechnology* 2011, **29**(7):644-U130.
47. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R, Genome Project Data Processing S: **The Sequence Alignment/Map format and SAMtools.** *Bioinformatics* 2009, **25**(16):2078-2079.
48. Katoh K, Standley DM: **MAFFT multiple sequence alignment software version 7: improvements in performance and usability.** *Mol Biol Evol* 2013, **30**(4):772-780.
49. Capella-Gutierrez S, Silla-Martinez JM, Gabaldon T: **trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses.** *Bioinformatics* 2009, **25**(15):1972-1973.
50. Darriba D, Taboada GL, Doallo R, Posada D: **ProtTest 3: fast selection of best-fit models of protein evolution.** *Bioinformatics* 2011, **27**(8):1164-1165.
51. Guindon S, Gascuel O: **A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood.** *Systematic Biology* 2003, **52**(5):696-704.

Table

Table 1. Data generated in this study and summary of the viruses

Rice samples	Data generated			viruses	
	No. reads	data yield (Gb)	contigs	Novel viruses	Known viruses
Rby1-21	70,923,042	12.69	328,146	RMV	RTV1
				RPeV	RRSV
				RPiV2	
				RNV	
Rby2-45	84,627,146	10.64	586,111	RMV	RTV1
				RPhV	RPiV1
				RPeV	RRSV
				RTV2	NLRV
				RPiV3	RToV
				RTV3	

Figures



Figure 1

Phenotypic analysis of the initial rice backcross generation. a Mature Ms55 (left), TZ21 (middle), and Rby1-21 (right) plants. Rby1-21 is shown as a representative rice plant from lines Rby1 and Rby2 infested by BPH; the rice plants grew in the soil for 18 weeks. Bar=30 cm; b Mature Rby1-N65. Rby1-N65 is shown as a representative rice plant from lines Rby1 and Rby2 not infested with BPH. Bar=30 cm; c Mature

panicles of TZ21 (left), Ms55 (middle) and Rby1-21 (right). Bar=10 cm; d Magnified image of a TZ21 panicle; e Magnified image of a Rby1-21 panicle.

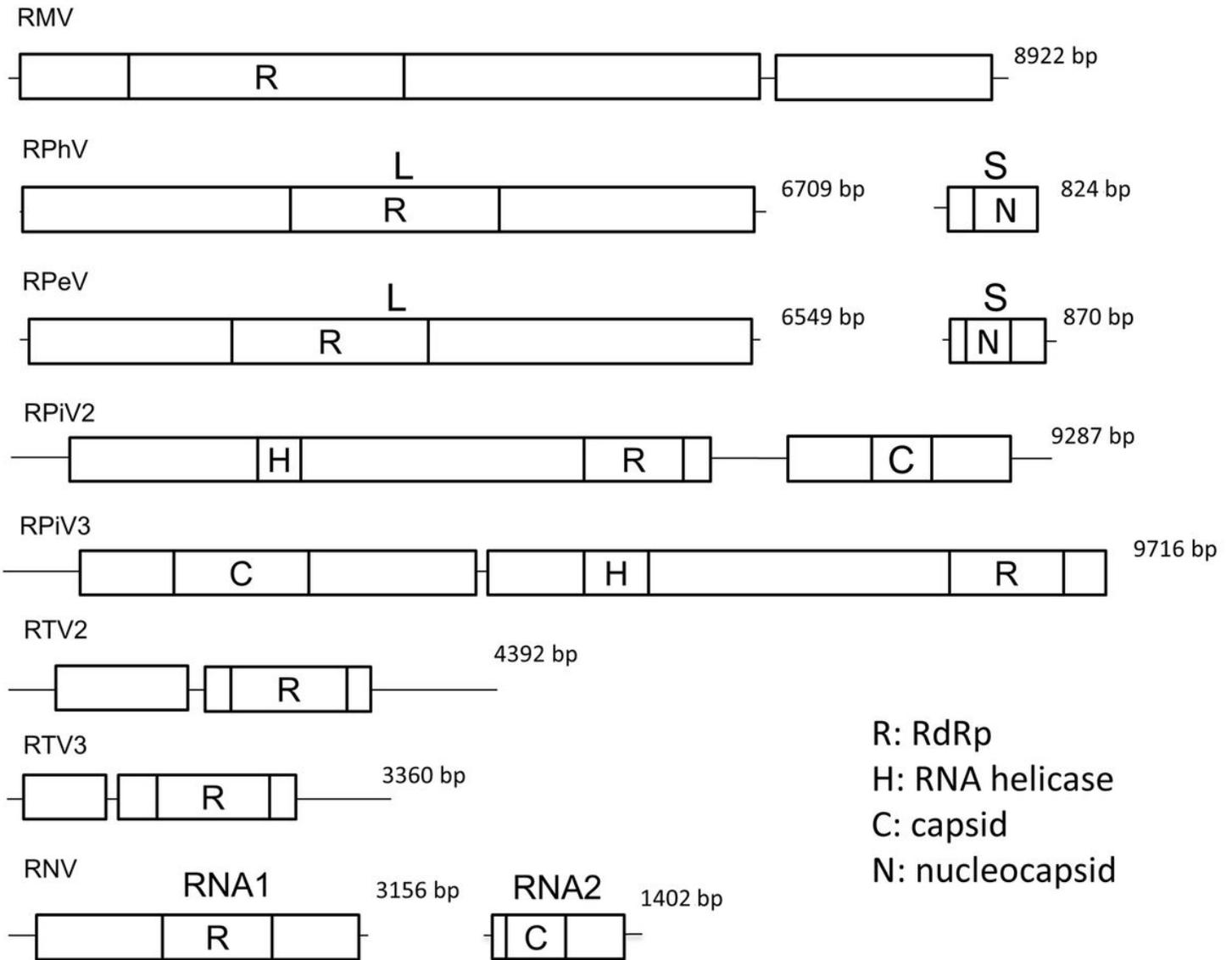
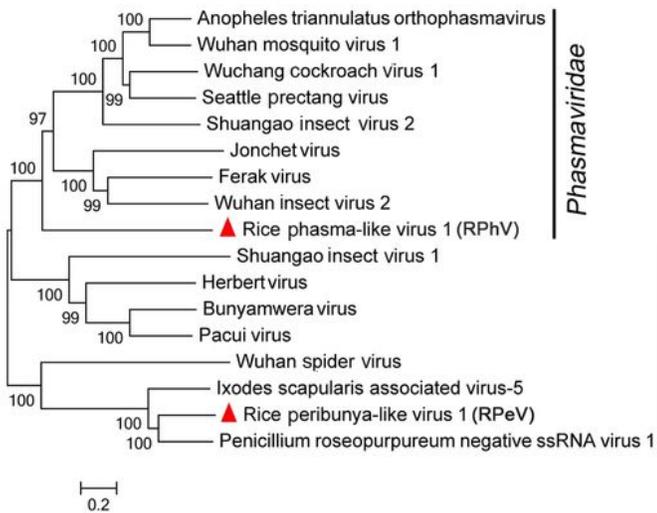


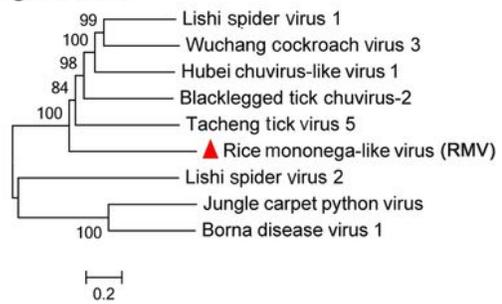
Figure 2

Genome structures of novel viruses. Predicted viral proteins homologous to known viral proteins are shown according to their putative functions. L: Large segment; S: small segment

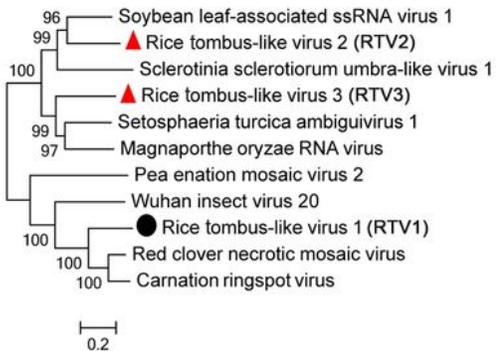
Bunyavirales



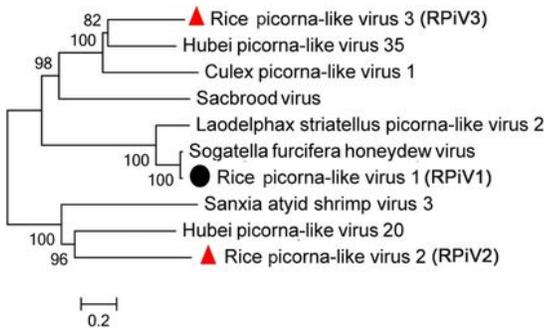
Mononegavirales



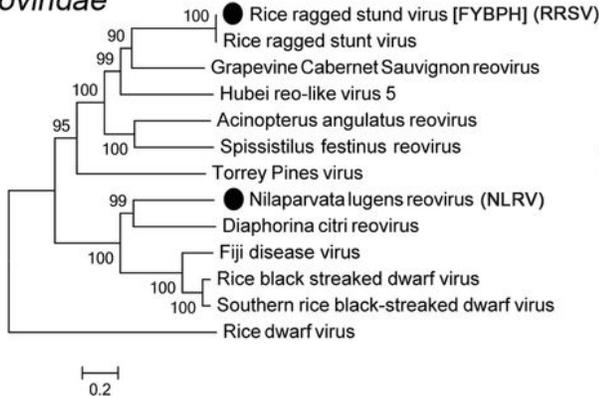
Tombusviridae



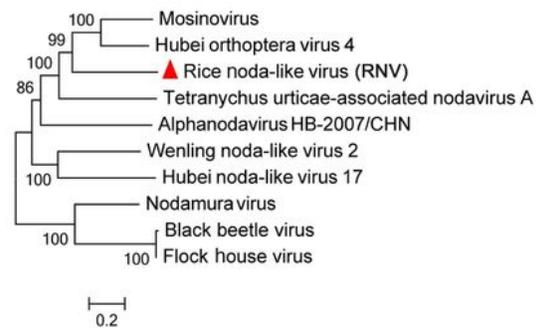
Picornaviridae



Reoviridae



Nodaviridae



Totiviridae

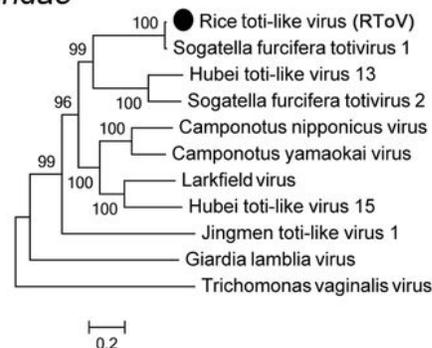


Figure 3

Evolutionary relationship of identified viruses in Rby1-21 and Rby2-45 rice plants. The maximum likelihood phylogenetic trees show the position of novel viruses (solid red triangles) and known viruses (solid black circles) in the context of representatives of their closest relatives.

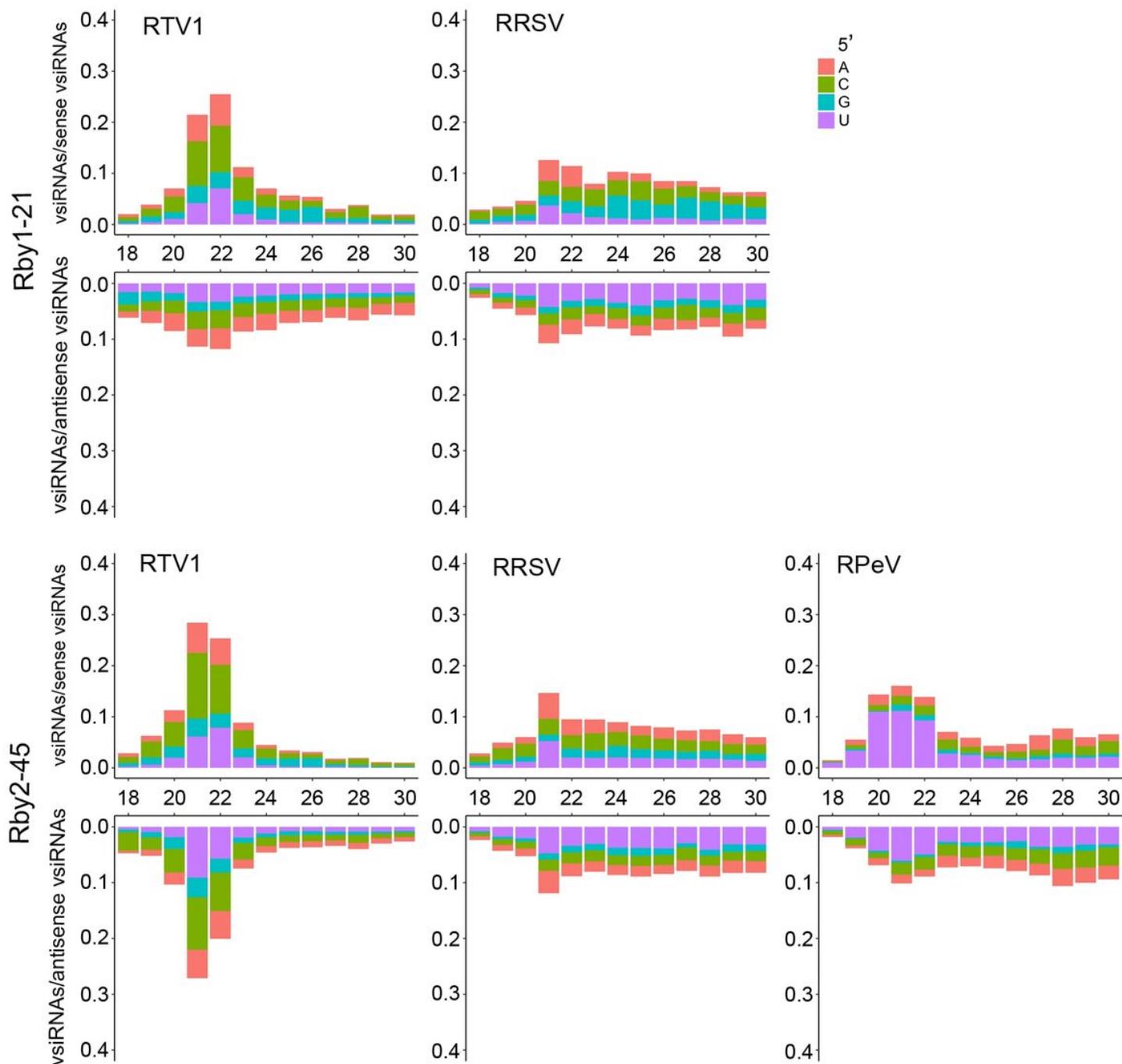


Figure 4

Small RNA analysis of RTV1 and RRSV in Rby1-21 as well as RTV1, RRSV and RPeV in Rby2-45. Size distribution, polarity and 5'-terminal nucleotide of virus-derived small RNAs derived from three viruses (RTV1, RRSV and RPeV). The relative abundance of different-sized sense vsiRNAs (top) is shown as the proportion of sense vsiRNAs, and that of different-sized antisense vsiRNAs (bottom) is presented as the proportion of antisense vsiRNAs. The bars are colored according to the 5'-terminal nucleotide compositions of vsiRNAs.

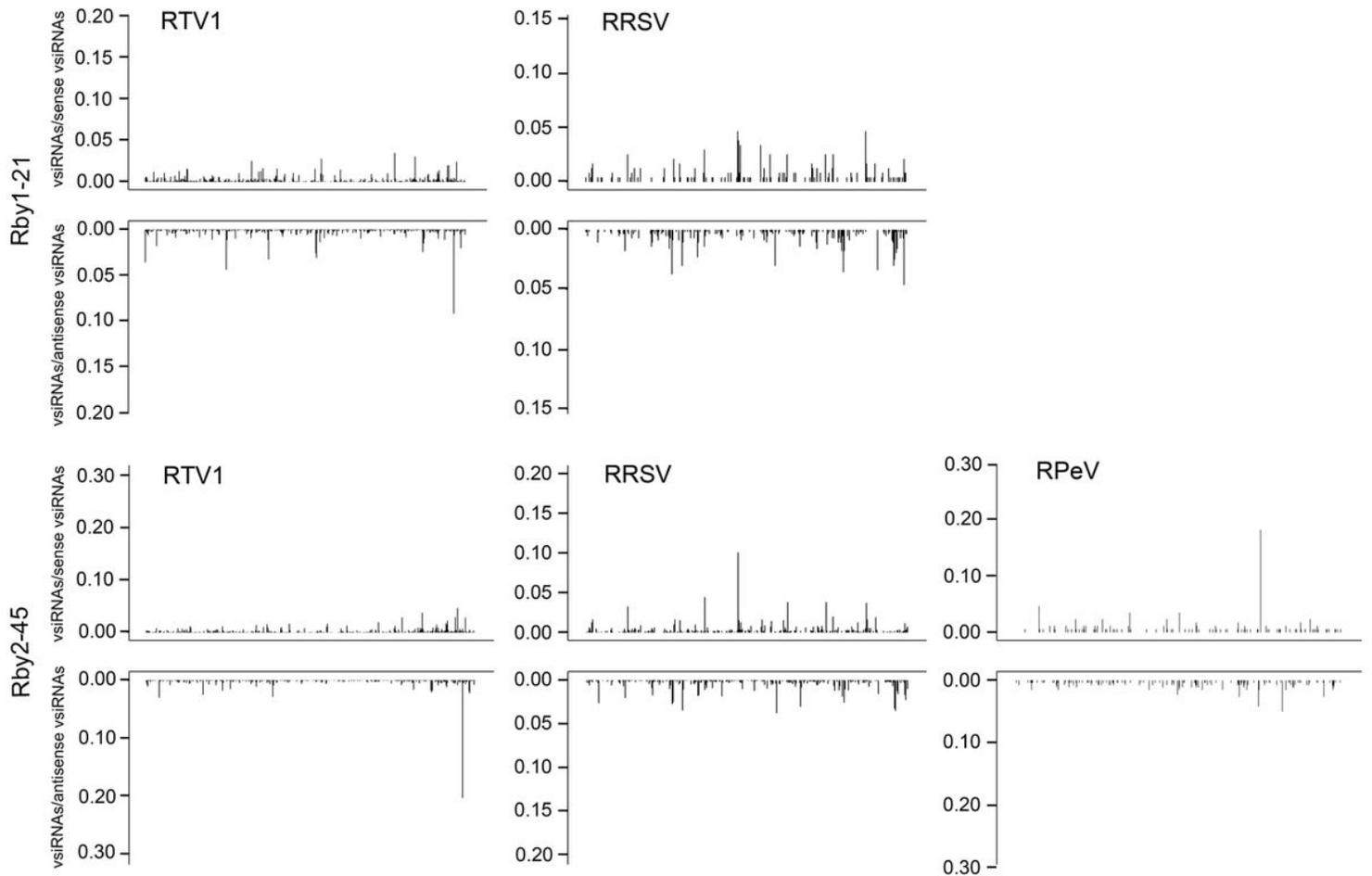


Figure 5

Highly abundant 21-nt vsiRNAs for targeting the viral genomes. Mapping of the 21 sense and antisense vsiRNAs to the full-length genomes of RTV1 and RRSV in Rby1-21 as well as RTV1, RRSV and RPeV in Rby2-45. The lines plotted above the x axis represent vsiRNAs that map to the positive strand, and those plotted below represent those that mapped to the negative strand.

Supplementary Files

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

- [FigureS3.tif](#)
- [FigureS2.tif](#)
- [FigureS3.tif](#)
- [Sl.zip](#)
- [SupplementaryTablesandFigures.pdf](#)
- [FigureS1.tif](#)
- [FigureS1.tif](#)

- [SupplementaryTablesandFigures.pdf](#)
- [SI.zip](#)
- [FigureS2.tif](#)