



A Novel Rice Curl Dwarf-Associated Picornavirus Encodes a 3C Serine Protease Recognizing Uncommon EPT/S Cleavage Sites

Tianze Zhang^{1†}, Chenyang Li^{1†}, Mengji Cao^{2†}, Dan Wang¹, Qi Wang¹, Yi Xie¹, Shibo Gao¹, Shuai Fu^{1*}, Xueping Zhou^{1,3*} and Jianxiang Wu^{1,4*}

OPEN ACCESS

Edited by:

Elvira Fiallo-Olivé,
Institute of Subtropical
and Mediterranean Horticulture La
Mayora, Spain

Reviewed by:

Marc Fuchs,
Cornell University, United States
Juan Jose Lopez-Moya,
Centre for Research in Agricultural
Genomics (CRAG), Spain

*Correspondence:

Shuai Fu
fushuai@zju.edu.cn
Xueping Zhou
zzhou@zju.edu.cn
Jianxiang Wu
wujx@zju.edu.cn

† These authors have contributed
equally to this work and share first
authorship

Specialty section:

This article was submitted to
Microbe and Virus Interactions with
Plants,
a section of the journal
Frontiers in Microbiology

Received: 12 August 2021

Accepted: 21 September 2021

Published: 13 October 2021

Citation:

Zhang T, Li C, Cao M, Wang D,
Wang Q, Xie Y, Gao S, Fu S, Zhou X
and Wu J (2021) A Novel Rice Curl
Dwarf-Associated Picornavirus
Encodes a 3C Serine Protease
Recognizing Uncommon EPT/S
Cleavage Sites.
Front. Microbiol. 12:757451.
doi: 10.3389/fmicb.2021.757451

¹ State Key Laboratory of Rice Biology, Institute of Biotechnology, Zhejiang University, Hangzhou, China, ² National Citrus Engineering and Technology Research Center, Citrus Research Institute, Southwest University, Beibei, China, ³ State Key Laboratory for Biology of Plant Diseases and Insect Pests, Institute of Plant Protection, Chinese Academy of Agricultural Sciences, Beijing, China, ⁴ Hainan Research Institute of Zhejiang University, Hainan, China

Picornaviruses cause diseases in a wide range of vertebrates, invertebrates and plants. Here, a novel picornavirus was identified by RNA-seq technology from rice plants showing dwarfing and curling symptoms, and the name rice curl dwarf-associated virus (RCDaV) is tentatively proposed. The RCDaV genome consists of an 8,987 nt positive-stranded RNA molecule, excluding a poly(A) tail, that encodes two large polyproteins. Using *in vitro* cleavage assays, we have identified that the RCDaV 3C protease (3Cpro) as a serine protease recognizes the conserved EPT/S cleavage site which differs from the classic Q(E)/G(S) sites cleaved by most picornaviral 3C chymotrypsin-like cysteine proteases. Therefore, we comprehensively deciphered the RCDaV genome organization and showed that the two polyproteins of RCDaV can be cleaved into 12 mature proteins. We found that seven unclassified picornaviruses also encode a 3Cpro similar to RCDaV, and use the highly conserved EPT/S as the cleavage site. The precise genome organizations of these viruses were illustrated. Moreover, RCDaV and the seven unclassified picornaviruses share high sequence identities and similar genome organizations, and cluster into a distinct clade in the order *Picornavirales*. Our study provides valuable information for the understanding of picornaviral 3Cpros, deciphers the genome organization of a few relatively obscure picornaviruses, and lays the foundation for further pathogenesis research on these viruses.

Keywords: RNA-seq, rice curl dwarf-associated virus, picornavirus, 3C protease, serine protease, cleavage site

INTRODUCTION

According to the 2020 taxonomic classification ratified by the International Committee on Taxonomy of Viruses (ICTV)¹, the order *Picornavirales* contains eight officially classified families: *Picornaviridae*, *Dicistroviridae*, *Iflaviridae*, *Marnaviridae*, *Polycipiviridae*, *Secoviridae*, *Caliciviridae*, and *Solinviviridae*, with one subfamily, 103 genera, three subgenera, and a total of 323 species

¹https://talk.ictvonline.org/ictv-reports/ictv_9th_report/positive-sense-rna-viruses-2011/w/posrna_viruses/227/picornavirales

as well as an unassigned group containing unclassified picornaviruses^{2,3}. Viruses in the order *Picornavirales* have positive sense RNA genomes with 5'-bound VPg (viral protein genome-linked) and 3'-poly(A) encapsidated in spherical viral particles about 30 nm in diameter. They infect vertebrates, arthropods, plants, fungi and algae (Sanfaçon et al., 2009). Their ORFs encode polyproteins, in which the non-structural region contains a typical "replication block" with a type III helicase, a chymotrypsin-like fold (cysteine) protease and a type I RNA-dependent RNA polymerase (Hel-Pro-Pol) (Le Gall et al., 2008).

Genome organizations vary significantly among picornaviruses. The genome of viruses in the family *Marnaviridae* or *Dicistroviridae* is dicistronic, while viruses in the families *Picornaviridae* and *Iflaviridae* possess a monocistronic genome, except the dicistronic viruses in the genus *Dicpivirus* (Yinda et al., 2017). Viruses in the family *Polycipiviridae* are polycistronic, with four or more consecutive ORFs in the 5'-proximal region and a long ORF encoding non-structural proteins in the 3'-proximal region (Olendraite et al., 2017). Viruses in the family *Secoviridae* possess either monopartite or bipartite genomes (Sanfaçon et al., 2009). It is noteworthy that plant-infecting picornaviruses are all classified in the family *Secoviridae* (Sanfaçon et al., 2009). Genomes of the viruses in the family *Caliciviridae* are not monocistronic but contain one or two additional ORFs (Le Gall et al., 2008). The capsid proteins of the viruses in the newly established family *Soliniviridae* can be expressed from subgenomic or genomic RNAs as an extension of the "replication block" (Brown et al., 2019). However, some picornaviruses have not been well classified yet as a result of divergent genome organizations, low sequence identities, as well as distant evolutionary relationships with viruses from the eight officially classified families, which are still referred to as unclassified picornaviruses.

Picornaviral polyproteins are post-translationally cleaved by virus-encoded 3C protease (3Cpro) and the leader protein (L) in a proteolytic manner, or co-translationally processed by the 2A protein via a non-proteolytic mechanism (Kjær and Belsham, 2018). The 3Cpro is responsible for most of the cleavage activities (Yang et al., 2017). Typical picornaviral 3Cpros are intermediates between the chymotrypsin-like serine proteases and the papain-like cysteine proteases (Sárkány and Polgár, 2003). Analysis of the crystal structure of picornaviral 3Cpro reveals its similarity with chymotrypsin-like serine protease (Allaire et al., 1994; Matthews et al., 1994), but a conserved cysteine (C) nucleophile replaces serine (S), resulting in a GxCG core motif instead of a GxSG core motif. Therefore, the classic catalytic triad in picornaviral 3Cpros contains a histidine (H), an aspartate/glutamate (D/E), and a conserved cysteine (Dessens and Lomonosoff, 1991; Sárkány and Polgár, 2003). However, the 3Cpros of the viruses in the family *Polycipiviridae* and several viruses in the family *Secoviridae* and *Marnaviridae* retain catalytic serine, indicating they encode the typical serine proteases (Olendraite et al., 2017; Mann and Sanfaçon, 2019).

Early studies have demonstrated the proteolytic activity of viral 3C protease using a series of *cis*- and *trans*-cleavage assays in rabbit reticulocyte cell-free transcription/translation system (Wetzel et al., 2013; Mann et al., 2017). In addition, the cleavage activity of 3Cpro encoded by *Ectropis obliqua* picorna-like virus (EOV) or rice tungro spherical virus (RTSV) has been identified using purified recombinant proteins expressed by *Escherichia coli* cells (Thole and Hull, 1998; Ye et al., 2012). The highly specific cleavage sites recognized by picornaviral 3Cpros are commonly Q/G, Q/S, and E/G dipeptides with specific amino acids (aa) flanking the cleavage sites (Seipelt et al., 1999; Sanfaçon, 2015), and it has been suggested that for the picornaviruses, the conserved His (H) residue in the S1 position of the substrate binding pocket (SBP) may be important for cleavage site recognition (Bazan and Fletterick, 1988; Mann and Sanfaçon, 2019). As 3Cpro plays a central role in polyprotein processing, classic picornaviral 3C cysteine proteases have been extensively studied in several families. However, research on the function and the cleavage sites recognition specificity of picornaviral serine 3Cpro remains scanty.

With the rapid development of high-throughput RNA-seq, a large number of novel viruses have been discovered in recent years (Massart et al., 2014; Shi et al., 2016). For example, Shi et al. (2016) discovered 1,445 novel RNA viruses in over 220 invertebrate species through deep transcriptome sequencing. In 2017, rice plants showing dwarfing and curled tillers were collected in paddy fields in the Zhejiang Province, China. Analysis of the sequencing data by RNA-seq allowed us to identify a new picornavirus, which was tentatively named rice curl dwarf-associated virus (RCDAV). Phylogenetic analysis showed that RCDAV and seven unclassified picornaviruses cluster into a distinct clade in the order *Picornavirales*, and encode a chymotrypsin-like serine protease with the conserved EPT/S cleavage sites.

MATERIALS AND METHODS

Virus Source and Electron Microscopy

Rice plants showing dwarfing and curling symptoms were collected in 2017 from rice fields in the Zhejiang Province, China. For transmission electron microscope (TEM), the collected samples were homogenized in sterile deionized water (1 g tissue/5 mL water). After 5 min centrifugation at 5,000 × g, the supernatant was examined under a TEM (JEOL JEM-1010, Tokyo, Japan).

RNA Sequencing and *de novo* Assembly

Total RNA was extracted from rice tissues using TRIzol reagent (Invitrogen, Carlsbad, United States). Total RNA (3 µg) was used for cDNA library constructions by the Zhejiang TianKe High-Technology Development Co., Ltd. (Hangzhou, China), and sequenced using the Illumina HiSeqTM 4000 sequencing system (Illumina, San Diego, United States) as described previously (Zhang et al., 2020). The raw sequencing data was processed after removing low quality reads using the CLC Genomics Workbench 9.5 (Qiagen, Valencia, United States). The resulting

²<http://www.ictvonline.org/>

³www.picornaviridae.com

high quality reads were then mapped to the rice genome⁴ and the reads matched the rice genome were removed. The non-rice reads were then imported to the Trinity program for *de novo* assembly (Grabherr et al., 2011), and the assembled contigs were then subjected to BLASTx and BLASTp searches in the NCBI databases. The identified virus-like sequences were extracted according to the annotation information.

RT-PCR, and 3' and 5' Rapid Amplification of the cDNA Ends-PCR

The 3' and 5' ends of the viral genomic RNA were obtained through rapid amplification of the cDNA ends (RACE) using a SMARTer RACE cDNA amplification kit as instructed (Clontech, Mountain View, United States). The resulting sequences were checked manually and then assembled to produce the final viral genome sequence. The complete viral genome sequence was confirmed by re-sequencing several fragments amplified by RT-PCR using specific primers (**Supplementary Table 1**). The conserved domains in the viral genome sequence were identified using the Conserved Domain Search Service (CD-Search) at the NCBI web server⁵ (Marchler-Bauer et al., 2015).

RT-PCR Detection

Total RNA was isolated from rice plants using TRIzol reagent. The cDNA was generated from total RNA by reverse transcription using ReverTra Ace qPCR RT Master Mix with gDNA Remover (TOYOBO, Osaka, Japan) following the recommended protocol. The PCR was set up using the Green Taq Mix (Vazyme, Nanjing, China) according to the manufacturer's instructions. Primer pair RCDaV-detection-F1/R1 (**Supplementary Table 1**) was used to detect RCDaV. PCR reaction (20 μ L each) contained 10 μ L 2 \times Green Taq mix (Vazyme), 1 μ L for each primer (10 μ mol/L each), 1 μ L cDNA and 7 μ L sterile deionized water. Thermal cycles settings were 94°C for 2 min; 30 cycles of 94°C for 30 s, 53°C for 30 s, and 72°C for 10 s; and the final extension was 72°C for 10 min.

Plasmid Construction

Protein expression vector pET-32a was modified by replacing the 6 \times His tags at C-terminus with a 3 \times FLAG tag. To construct a pET-32a-3 \times FLAG vector, the original pET-32a vector was linearized via double digestion with *NotI* and *XhoI* restriction enzymes (Thermo Fisher Scientific, Waltham, United States). A 3 \times FLAG tag containing a stop codon was fused to vector using the ClonExpress II one step cloning kit (Vazyme).

The substrate segments were PCR-amplified, and purified using a DNA gel extraction kit (Corning Life Sciences, Lowell, United States). Vector pET-32a-3 \times FLAG was linearized through double digestion with *BamHI* and *Sall* restriction enzymes (Thermo Fisher Scientific). The purified PCR product and the linearized vector were fused using ClonExpress II one step cloning kit (Vazyme) according to the manufacturer's instructions.

The protease segments (RCDaV-Pro_{1054–1496}, MaPV-Pro_{1094–1490}, ApGIV1-Pro_{956–1355}), MBP tag, and SUMO tag with a stop codon were PCR-amplified. The original pET-28a vector was linearized through backward PCR-amplification to remove its tags, except the C-terminal His tag. Then, the purified PCR products were fused with the linearized vector using ClonExpress MultiS one step cloning kit (Vazyme). The plasmid for expressing tag-free 3Cpro was generated through fusing PCR-amplified 3Cpro with the *NcoI/XhoI* linearized pET-28a vector using the ClonExpress II one step cloning kit (Vazyme). Primers used in this section are listed in **Supplementary Table 1**, details of the constructs are listed in **Supplementary Table 2**.

Site-Directed Mutagenesis

The site-directed mutations were introduced to the gene segments by using Mut Express II Fast Mutagenesis Kit (Vazyme) according to the manufacturer's instructions. The plasmids containing substrates or 3Cpros were backward PCR-amplified using primers carrying specific mutations of the codons, and then the purified PCR product was self-ligated. Primers used in this section are also listed in **Supplementary Table 1**, details of the mutated constructs are listed in **Supplementary Table 2**.

In vitro Cleavage Assays

Substrates and 3Cpro proteins were prepared using a modified *E. coli* cell extract-based cell free protein expression kit (GZL Bioscience, Hangzhou, China). Briefly, the pET-32a and pET-28a vectors with desired gene sequences were individually amplified through PCR to generate the DNA templates. The resulting templates (15 μ L/template) was used in a 35 μ L reaction mixture composed of 17 μ L reaction buffer, 13 μ L *E. coli* cell extract, and 5 μ L ddH₂O. The mixtures were incubated at 30°C for 3 h for protein expression.

For *in vitro cis*-cleavage assay, after 3 h protein expression, the incubated reaction mixtures were individually mixed with 2 \times SDS-PAGE sample buffer, boiled for 5 min, and then subjected to western blot analysis. For *in vitro trans*-cleavage assay, the protein expression incubation time was reduced to 1.5 h at 30°C. The resulting substrate and 3Cpro were mixed, and then incubated at 28°C for 1.5 h followed by western blot analysis.

Prokaryotic Protein Expression and N-Terminal Edman Degradation Sequencing

FLAG-tagged fusion proteins were expressed with pET32a-3 \times FLAG vector in *E. coli* strain BL21 (DE3) and purified with FLAG-beads followed by 12% SDS-PAGE. The cleavage protein bands were cut out and subjected to N-terminal Edman degradation sequencing by Tailian biotech Co., Ltd. (Beijing, China).

Western Blot

Western blot analyses were conducted as previously described (Fu et al., 2018) with an anti-FLAG or anti-6 \times His murine antibody.

⁴<http://rice.plantbiology.msu.edu/>

⁵<https://www.ncbi.nlm.nih.gov/Structure/cdd/wrpsb.cgi>

Phylogenetic Analyses

Sequences of viruses in different families in the order *Picornavirales* and the unassigned members were retrieved from the GenBank database (**Supplementary Table 3**). The deduced amino acid sequences of RdRPs and 3Cpros were aligned using the Muscle v 3.8.31 (Edgar, 2004) and the phylogeny analyses were performed using the MEGA X software (Kumar et al., 2018) via the maximum-likelihood method based on the JTT matrix-based model with a bootstrap of 1000 replications (Felsenstein, 1985; Jones et al., 1992). The resulting phylogeny trees were presented using the iTOL online tool⁶ (Letunic and Bork, 2019).

Cleavage Site Conservation Analyses and Sequence Logo Generation

The sequence of cleavage sites (P4–P4′) in RCDaV and the seven unclassified members were extracted and analyzed using the TBtools software (Chen et al., 2020). The sequence logos were generated with the same software to demonstrate the conservation of cleavage sites.

Gene Synthesis

MaPV-P2_{1–310} (5,033–5,962 nt) and *MaPV-Pro_{1094–1490}* (1,493–2,683 nt) of maize-associated picornavirus (MaPV, accession number MF425855) and *ApGIV1-P2_{1–199}* (3,196–3,792 nt) and *ApGIV1-Pro_{956–1355}* (6,061–7,260 nt) of aphid glycines virus 1 (ApGIV1, accession number KM015260) were synthesized by the GenScript Biotech (Nanjing, China).

Data Availability

The genome sequence of RCDaV has been deposited in GenBank as accession number MW725267.

RESULTS

Identification of a Novel Rice-Associated Picornavirus

During a field survey in the Zhejiang Province, China in August 2017, rice plants showing dwarfing and curling symptoms were observed and sampled (**Figures 1A,B**). Known rice-infecting viruses (i.e., rice black-streaked dwarf virus, southern rice black-streaked dwarf virus, rice stripe virus, rice ragged stunt virus, rice dwarf virus, rice gall dwarf virus, rice grassy stunt virus, rice stripe mosaic virus) failed to be detected in these rice plants by Dot-ELISA or RT-PCR (data not shown). However, electron microscopy analysis showed the presence of non-enveloped spherical particles with a diameter of approximately 30 nm in the rice plant crude extract (**Figure 1C**). To investigate the nature of these virus-like particles, we extracted total RNA from collected rice samples and analyzed them through RNA-seq. The RNA-seq analysis produced a total of 86,216,078 clean reads, and 77,020,091 of them (89.33%) were mapped to the reference rice genome and thus removed. A total of 9,195,987 clean reads

were assembled to produce 63,682 contigs with 200–14,135 nt in length. Among these reads, 118,000 were further assembled into a long contig of 8,675 nt, and the BLASTx result revealed that this contig may represent a novel picornavirus. Then, the full-length sequence of this putative viral RNA genome was obtained by RT-PCR, 5′ and 3′UTR RACEs, revealing an 8,987 nt positive-stranded RNA genome, excluding a poly(A) tail (**Figure 1D**). This viral genome contains two large ORFs: a 2,409 nt ORF encodes an 802 aa protein (thereafter referred to as ORF1) and a 5,682 nt ORF encodes an 1,893 aa protein (ORF2). These two ORFs are separated by a 297 nt intergenic region (IGR). The complete genome sequence has been submitted to GenBank under the accession number MW725267.

Using the NCBI Conserved Domain Database (CDD) tool, we identified three conserved capsid protein (CP) domains in ORF1, including two rhv-like domains (cd00205, aa 56–235, and aa 343–529) and a cricket paralysis virus (CrPV) capsid-like domain (cl07393, aa 636–769) (**Figure 1D**). The conserved non-structural protein domains in ORF2 include an RNA helicase domain (pfam00910, aa 678–788), a Peptidase_C3G domain (cl13774, aa 1088–1310), and an RNA_dep_RNAP domain (cd01699, aa 1,507–1,804) (**Figure 1D**). The BLASTp search reveals that the RNA_dep_RNAP domain shares the closest aa sequence identity (90.6%) with the polyprotein of maize-associated picornavirus (MaPV) (AUH27292.1) and 65–90% aa sequence identities with several unclassified picornaviruses such as *Tetranychus urticae*-associated picorna-like virus 1 (TUaPV1) (QIN54759.1), aphid glycines virus 1 (ApGIV1) (AHC72013.1) and cherry virus Trakiya (CVT) (YP_009551963.1) (**Supplementary Table 4**). According to these results, we propose this virus as a novel species in the order *Picornavirales*, and tentatively name it rice curl dwarf-associated virus (RCDaV). Analysis of the sampled rice plants through RT-PCR showed that RCDaV was present in rice leaves, stems, and roots (**Supplementary Figure 1**). RCDaV was also detected in barnyard grass plants with similar dwarfing and curling symptoms, which were collected in the same rice field (**Supplementary Figure 1**).

The genome of most picornaviruses is a single-stranded positive-sense RNA, and lacks a 5′ cap structure needed for the initiation of protein synthesis. Instead, a small viral protein 3B (also known as VPg) is covalently linked to the 5′ end of the viral genomic RNA (Le Gall et al., 2008; Sanfaçon et al., 2009). Thus, we postulate that RCDaV also follows this rule, like other picornaviruses. Moreover, the predicted AUG translation initiation codons for the two polyproteins are set in Kozak consensus sequences (AAAA₅₀₁UGG, ATCA₃₂₀₇UGG) (Kozak, 1999).

In this study, we adopted the L344 nomenclature system (Rueckert and Wimmer, 1984) to name RCDaV proteins because the arrangement of RCDaV proteins is similar to that of canine picodistrovirus (JN819202) in the family *Picornaviridae* (Woo et al., 2012). As shown in **Figure 1D**, RCDaV polyproteins are divided into three different regions: region P1 encodes viral CPs, region P2 encodes protein 2AB and 2C, and region P3 encodes proteins 3A–3D known as viral protein-processing and genome-replication proteins. According to the conserved domain identified in the two polyproteins of RCDaV, at least six putative

⁶<https://itol.embl.de>

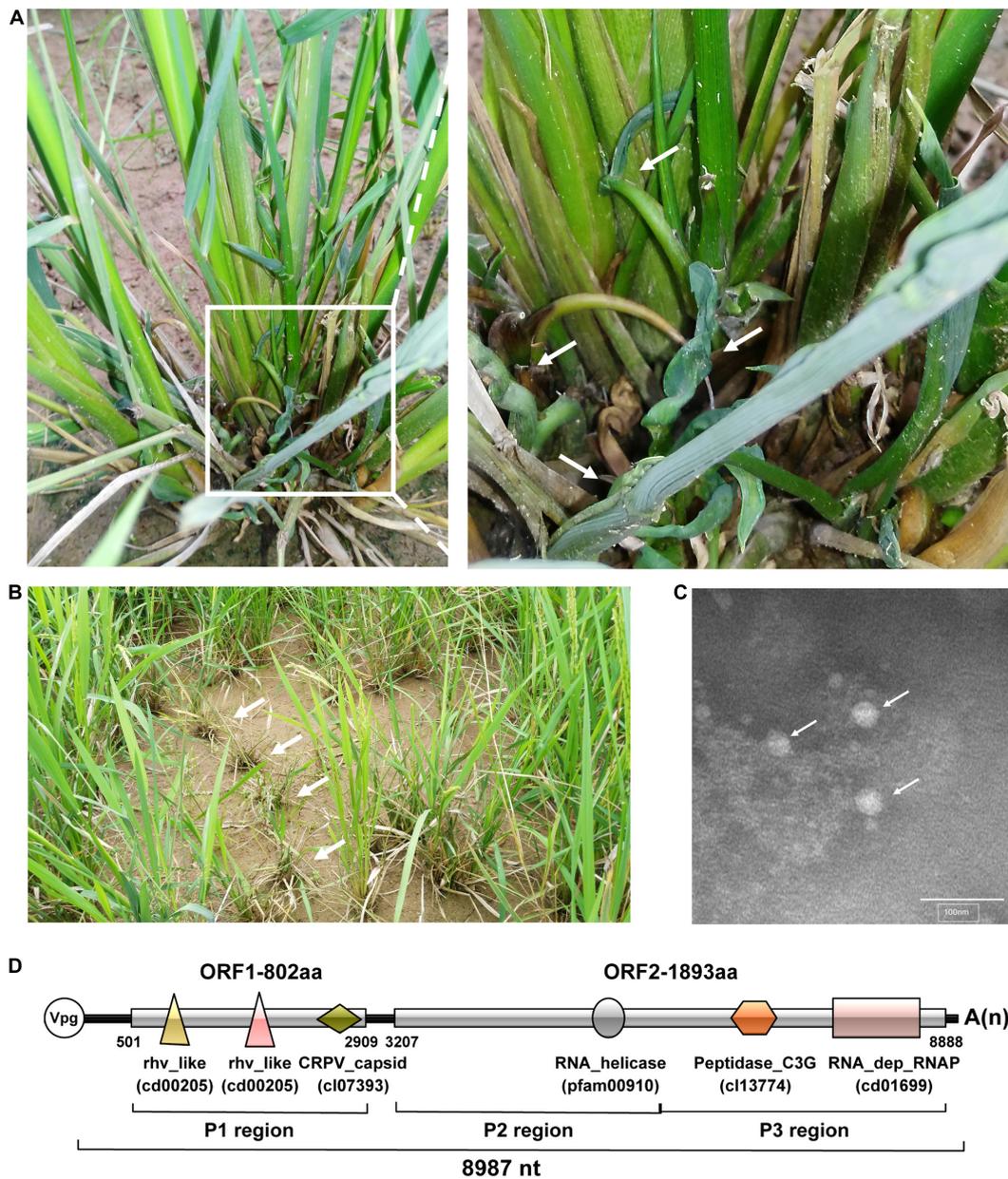


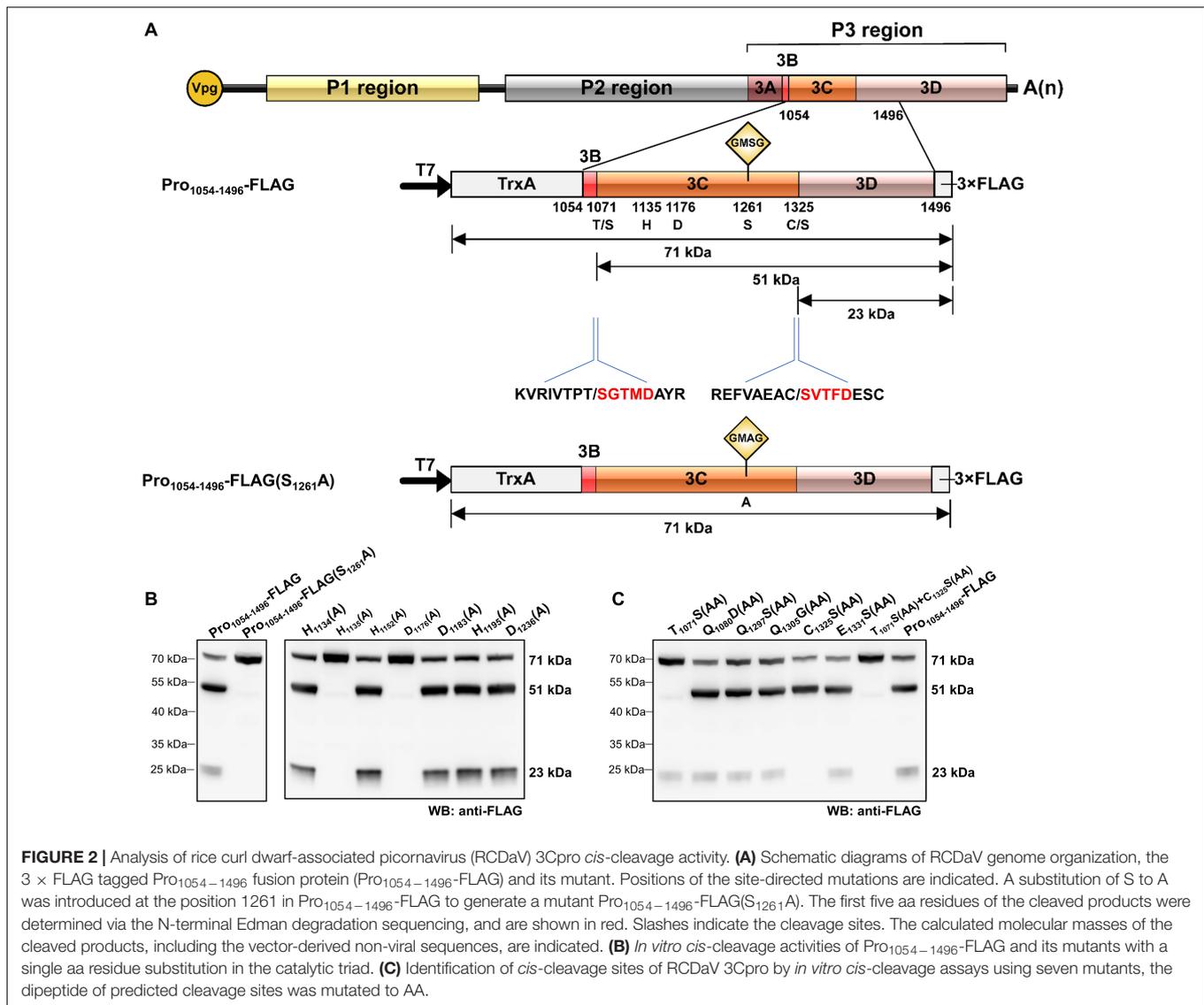
FIGURE 1 | Rice plant symptoms, virions, and genome organization of rice curl dwarf-associated picornavirus (RCDaV). **(A)** The rice plants showing curling and dwarfing symptoms. The right image is the enlargement of the box in the left image. White arrows indicate curled tillers. **(B)** The rice plants showing dwarfing in the field. White arrows indicate symptomatic rice plants. **(C)** Electron micrograph of putative RCDaV virions. Extracts were prepared from curled tillers of rice and examined under a transmission electron microscope after negative staining with a 2% phosphotungstic acid solution, pH 7.0. Bar = 100 nm. White arrows indicate viral particle-like structures. **(D)** A schematic diagram of RCDaV genome organization. Three protein regions as well as six predicted domains are shown below the diagram. The six different domains are represented by different shapes and colors, and the name of the domains are marked below the shapes. The VPg is a viral protein covalently linked to the 5' end of the viral genomic RNA.

proteins were predicted, however, the boundaries and the precise cleavage sites of the structural and non-structural proteins need to be determined.

Cis-Cleavage Activity of RCDaV 3Cpro

The picornaviral genome is generally translated into large precursor polyproteins followed by proteolytic processing

through self-released 3Cpro, which is a vital step for viral structural and non-structural proteins maturation. To determine the *cis*-cleavage activity of RCDaV 3Cpro, a gene segment containing 3Cpro domain (aa position 1,054–1,496 of polyprotein 2) was cloned into the modified pET32a vector to produce pET32a-Pro_{1054–1496}-3 × FLAG (referred to as Pro_{1054–1496}-FLAG thereafter) (Figure 2A). After 3 h expression



in the *Escherichia coli* cell extract-based transcription/translation system (*E. coli* cell-free system), the expressed proteins were analyzed using western blot analysis with anti-FLAG antibody. The result showed that a 71 kDa protein band, the expected size of the intact Pro₁₀₅₄₋₁₄₉₆-FLAG, as well as a 51 and a 23 kDa protein bands were detected (Figure 2B, lane 1). The presence of 71 kDa band indicated that the *cis*-cleavage was incomplete. We postulated that the 51 kDa protein band was the product autocatalytically cleaved from the N-terminus of 3Cpro and the 23 kDa protein band was cleaved from the C-terminus of 3Cpro. The intensity of the 23 kDa protein band was much weaker than that of the 51 kDa protein band, suggesting the *cis*-cleavage at 3Cpro N-terminus is more efficient than that at the C-terminus.

The catalytic triad of 3Cpro is a set of three coordinated amino acids in the active site of the enzyme. Each of these three key amino acids plays an essential role in the cleaving ability of the protease, and mutating these amino acids dramatically impacts 3Cpro cleavage activity (Dessens and Lomonosoff, 1991). To

exclude the possibility that the protein bands described above were cleaved by bacterial-derived proteases and to identify the catalytic triad of RCDaV 3Cpro, aa residues involved in the proteolytic process were predicted through comparing the sequences and 3D structures of 3Cpros from RCDaV and hepatitis A virus (HAV) (Supplementary Figures 2A,B). Then the predicted catalytic triad was determined via site-directed mutagenesis analyses. The result showed that the cleavage bands of 51 kDa and 23 kDa vanished in the H₁₁₃₅(A), D₁₁₇₆(A) and S₁₂₆₁(A) mutants of Pro₁₀₅₄₋₁₄₉₆-FLAG (Figure 2B). In classic picornaviral 3Cpro, the last amino acid of the catalytic triad is Cys (C), which is also the central amino acid of the core motif GxCG (Flint and Ryan, 1997). Interestingly, this Cys is replaced by Ser (S) in RCDaV 3Cpro, which results in a serine protease core motif GMSG (Figure 2B, lane 2).

To determine the exact *cis*-cleavage sites, we expressed Pro₁₀₅₄₋₁₄₉₆-FLAG in *E. coli* BL21 (DE3) cells and purified it using FLAG beads. After being separated by SDS-PAGE, the

51 and 23 kDa products were subjected to N-terminal Edman sequencing, respectively. The results revealed that the five amino acids at the N-terminus of the 51 kDa product were SGTMD (aa position 1,072–1,076 of the polyprotein 2), suggesting the cleavage occurred at the T₁₀₇₁/S dipeptide (**Figure 2A**). The five amino acids at the N-terminus of the 23 kDa product were SVTFD (aa position 1,326–1,330 of the polyprotein 2), suggesting that the cleavage occurred at the C₁₃₂₅/S dipeptide (**Figure 2A**). To validate this result, we mutated the T₁₀₇₁S, C₁₃₂₅S dipeptide, or putative cleavage site QG, ES dipeptides to AA, and *cis*-cleavage analyses demonstrated that the 51 kDa product was not detected in the T₁₀₇₁S(AA) mutant and the 23 kDa product was not detected in the C₁₃₂₅S(AA) mutant (**Figure 2C**, lanes 1, 5). As expected, other mutants gave the same cleaved products as Pro_{1054–1496}-FLAG (**Figure 2C**, lanes 2–4, 6, 8). In addition, the *cis*-cleavage activity of the double mutant [(T₁₀₇₁S(AA) + C₁₃₂₅S(AA))] was abolished (**Figure 2C**, lane 7). Taken together, these results indicate that the N- and C-terminal boundaries of the RCDaV 3Cpro are aa 1,072 and 1,325, and the cleavage occur at the T₁₀₇₁S and C₁₃₂₅S dipeptides.

Trans-Cleavage Activity of RCDaV 3Cpro

Using the preliminary experiments, we determined that the optimum temperature for protein expression was 30°C and the optimum temperature for cleavage was 28°C. In order to optimize the cleavage efficiency, the reaction temperature was set at 30°C for the first 1.5 h and 28°C for the last 1.5 h in the *trans*-cleavage experiment. To test the *trans*-cleavage activity of RCDaV 3Cpro, pET28a-MBP-Pro_{1054–1496}-SUMO-His (MBP-Pro_{1054–1496}-SUMO-His) and pET28a-MBP-Pro_{1054–1496}-SUMO-His(S₁₂₆₁A) were constructed and expressed in the *E. coli* cell-free system. The expressed MBP-Pro_{1054–1496}-SUMO-His and MBP-Pro_{1054–1496}-SUMO-His(S₁₂₆₁A) were used as the functional and the non-functional proteases, respectively, and the *cis*-cleavage-defective mutant Pro_{1054–1496}-FLAG(S₁₂₆₁A) was used as the substrate (**Figure 3A**). Western blot analyses showed that MBP-Pro_{1054–1496}-SUMO-His (protease) efficiently cleaved Pro_{1054–1496}-FLAG(S₁₂₆₁A) (substrate) to produce 51 kDa and 23 kDa cleavage products (**Figure 3B**, lane 1), while the substrate incubated alone remained intact (**Figure 3B**, lane 3). As expected, the substrate incubated with MBP-Pro_{1054–1496}-SUMO-His(S₁₂₆₁A) (inactive Protease) did not yield the cleaved protein bands (**Figure 3B**, lane 2), indicating that RCDaV 3Cpro is enzymatically active and able to mediate *trans*-cleavage.

To compare the *trans*- and *cis*- cleavage sites, we introduced the mutations into the substrate and then incubated them individually with MBP-Pro_{1054–1496}-SUMO-His. Western blot analyses showed that the *trans*-cleavage sites also occurred at the T₁₀₇₁/S and C₁₃₂₅/S dipeptides (**Figure 3C**). To optimize the *trans*-cleavage efficiency, we also analyzed the enzymatic activity of tag-free RCDaV 3Cpro_{1072–1325} (**Supplementary Figure 3A**). The result showed the tag-free 3Cpro_{1072–1325} could also catalyze the cleavage of the substrates used in this study (**Supplementary Figure 3B**, lane 1). Since the *trans*-cleavage activity of tag-free 3Cpro_{1072–1325} was relatively higher than

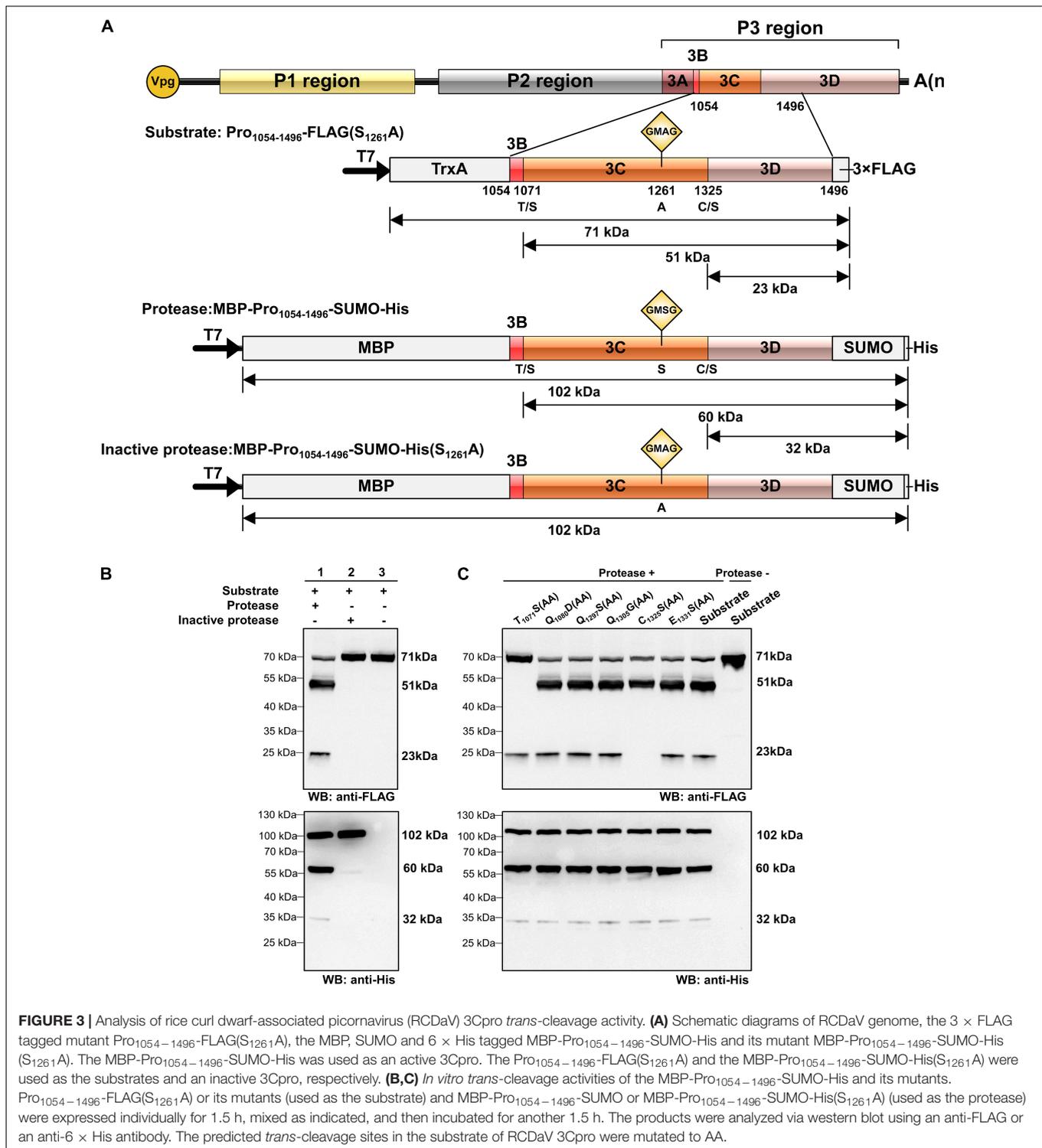
MBP-Pro_{1054–1496}-SUMO-His (**Supplementary Figure 3B**), we used the tag-free 3Cpro_{1072–1325} in the subsequent assays. Based on these results, RCDaV 3Cpro is identified as a serine protease with both *cis*-cleavage and *trans*-cleavage activity.

Precisely Mapping the Cleavage Sites on RCDaV Polyproteins

Polyproteins of picornaviruses are processed co- and post-translationally into 10–12 mature proteins (Lange et al., 2014). To determine the precise boundaries of functional viral proteins proteolytically cleaved by RCDaV 3Cpro, we analyzed the cleavage sites in RCDaV polyproteins via the *trans*-cleavage assays. We selected gene segments containing the boundaries of the predicted viral proteins to generate their 3 × FLAG-tagged fusion proteins as the substrates and used the tag-free 3Cpro_{1072–1325} as the protease (**Figure 4A**). After confirming the *trans*-cleavage of the substrates, mutations (e.g., QG-AA, ES-AA, TS-AA, and CS-AA) were introduced individually into the substrates through site-directed mutagenesis (**Figures 4B–H** and **Supplementary Table 2**). The cleavage sites were then determined by comparing the cleaved mutant substrates with the wild type substrates.

Some picornaviruses possess leader protein preceding the P1 region, such as the genera *Aphthovirus* and *Cardiovirus* in the family *Picornaviridae* (Devaney et al., 1988; Chen et al., 1995). The leader proteins of aphthoviruses are identified as papain-like cysteine proteinases that are able to self-cleave carboxy terminally (Devaney et al., 1988). In *trans*-cleavage assays, the substrate P1-1 (aa position 1–240 of the P1 region) produced a single 49 kDa protein band, in the presence or absence of 3Cpro_{1072–1325}, indicating that no *trans*-cleavage catalyzed by RCDaV 3Cpro or *cis*-cleavage catalyzed by L protein occurred (**Figure 4B**). Using substrate P1-2 (aa position 225–355 of the P1 region), we found that besides the intact 36 kDa protein band, an additional 15 kDa cleavage protein band catalyzed by 3Cpro_{1072–1325} was observed, except the T₂₄₄S(AA) mutant resulted in a complete inhibition of cleavage processing (**Figure 4C**). Besides the 39 kDa substrate P1-3 (aa position 505–665 of the P1 region), a 17 kDa cleavage protein catalyzed by 3Cpro_{1072–1325} was produced except the T₅₄₃S(AA) mutant (**Figure 4D**), suggesting that the T₅₄₃S dipeptide in P1-3 is important for the recognition by 3Cpro_{1072–1325}. Analyses of the P1 region revealed the presence of two cleavage sites (T₂₄₄/S and T₅₄₃/S), thus the P1 region can be processed in three CPs, but L protein is absent.

Further analyses of the RCDaV P2 region using *trans*-cleavage assays showed that this region can be cleaved into five proteins (**Figures 4E–G**). The *trans*-cleavage occurred at five highly conserved T/S dipeptides (aa position 132–133, 202–203, 368–369, 472–473, and 919–920, respectively), but not at the C/S or the Q(E)/G(S) dipeptides (**Figures 4E–G**). The cleavage at the T/S dipeptides in P2 region produce five proteins: a 15 kDa protein (referred to as X1), an 8 kDa protein (X2), a 20 kDa protein (X3), a 12 kDa protein (X4), and a 50 kDa 2C protein. We tentatively named the first four proteins as X1–X4 because they do not possess any motifs known in the classic picornaviral 2A or 2B proteins (Tseng and Tsai, 2007; Boros et al., 2014), and have no



sequence similarities to those identified in other picornaviruses. In addition, unexpected protein bands, indicated with asterisks, were found in the absence of RCDaV 3Cpro (**Figures 4E–H** and **Supplementary Figures 4A,B**). We postulate that these protein bands might be the products of endogenous proteases in the cell-free system. The 2C protein was predicted as a helicase, one of

the most conserved proteins encoded by viruses in the order *Picornavirales*. Picornaviral 2C proteins are all superfamily III helicases and contain three conserved motifs: GxxGxGK(S/T) (motif A), WWWxxDD (motif B), and KGx₄Sx₅(S/T)(S/T)N (motif C) (Hales et al., 2008). The motif A of RCDaV 2C is G₆₈₀LAGTRKS and its third “G” is substituted by “R” at the

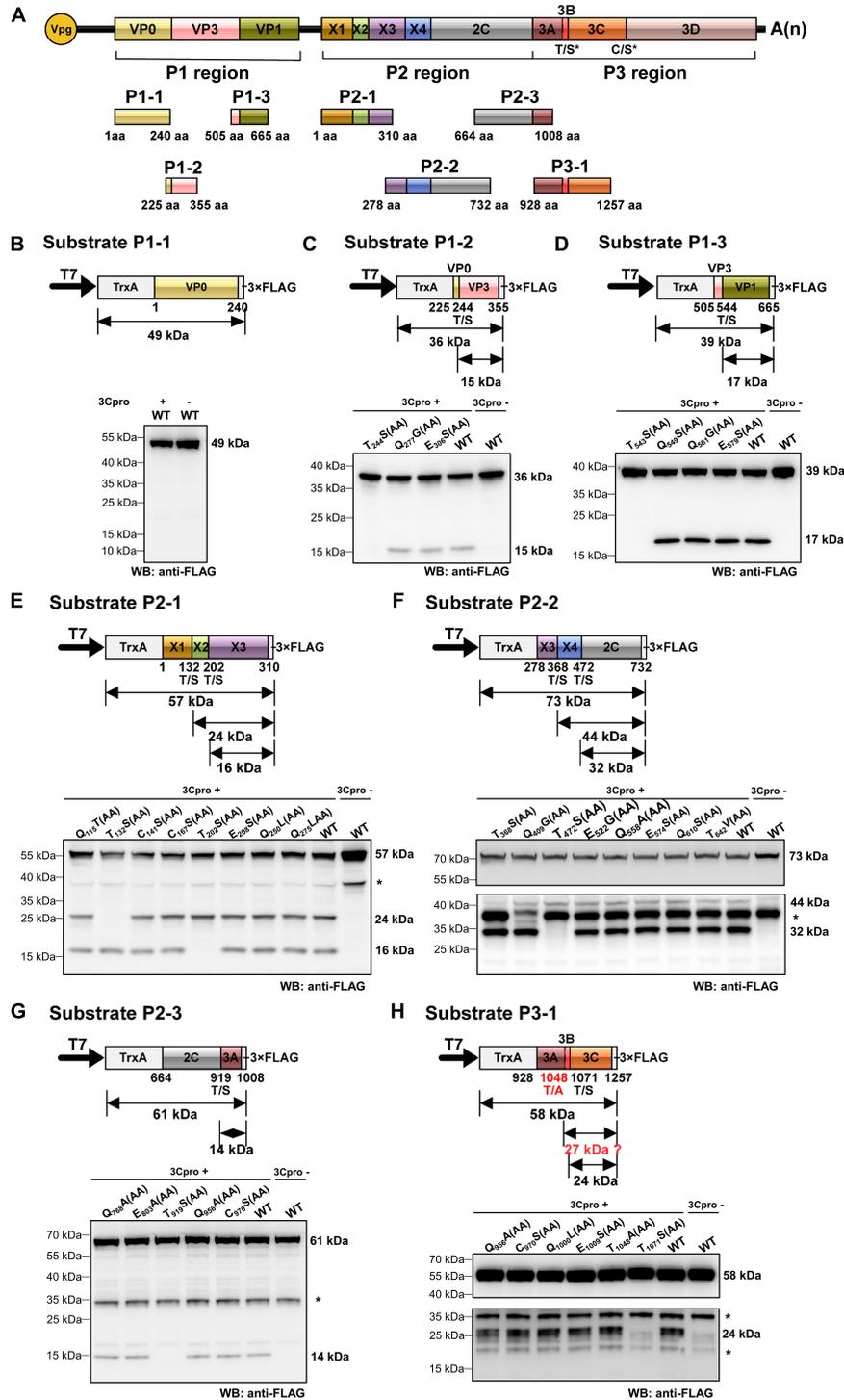
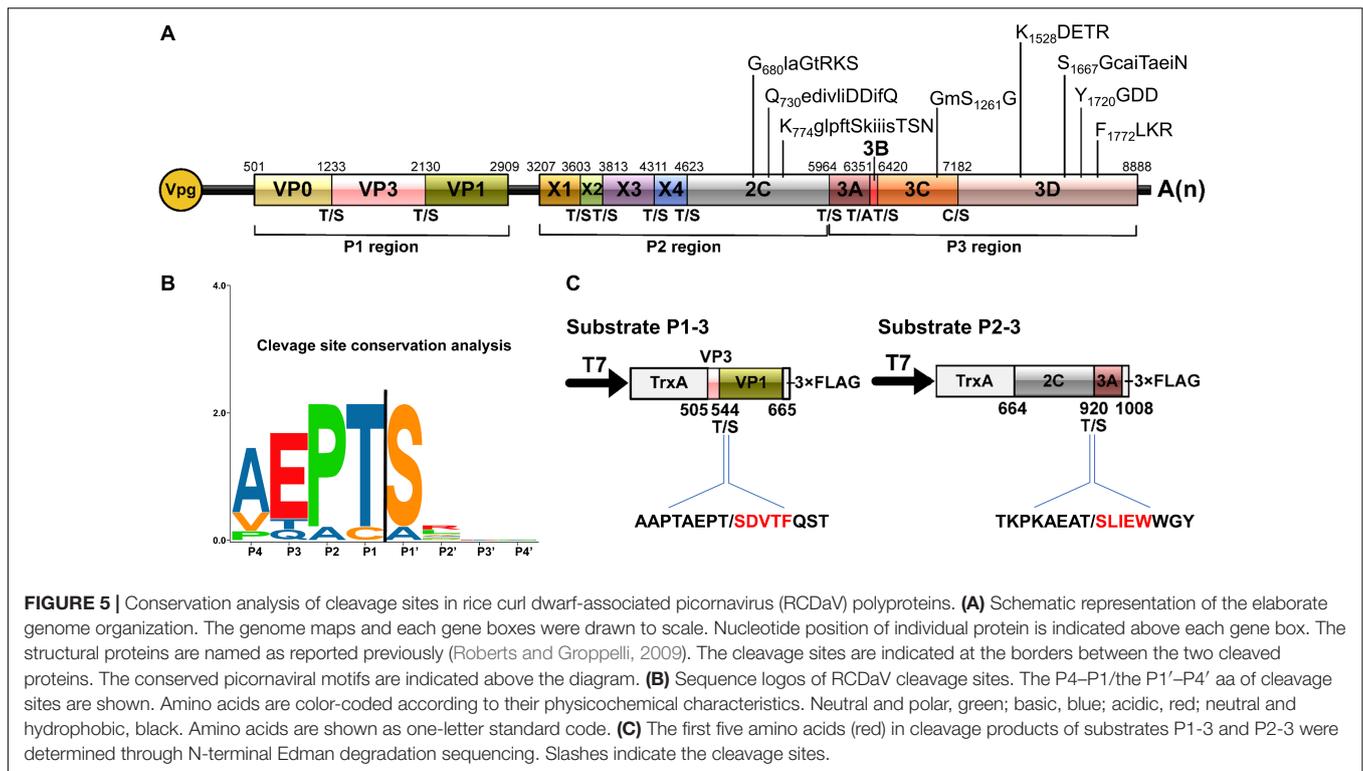


FIGURE 4 | Identification of cleavage sites in rice curl dwarf-associated picornavirus (RCDAV) polyproteins using *in vitro trans*-cleavage assay. **(A)** Schematic representation of protein segments from the three regions of RCDAV. **(B–D)** *In vitro trans*-cleavage analyses using the P1-1, P1-2, and P1-3 protein segments from P1 region as the substrates. **(E–G)** *In vitro trans*-cleavage analyses using the P2-1, P2-2, and P2-3 protein segments from P2 region as the substrates. **(H)** *In vitro trans*-cleavage analysis using the P3-1 protein segment from P3 region as the substrate. Site-directed mutants of the predicted cleavage sites were also used in individual assays. The calculated molecular mass of each protein segment is shown on the right side of the figure. The asterisks in **(E–H)** indicate non-specific protein bands. The lower panel blots in **(F,H)** were exposed for longer time to show the bands. Predicted T/A cleavage site is shown with a dashed line in **(H)**, since processing was not detected at this site in the *in vitro* assays. Numbers below the boxes indicate amino acids positions.



aa position 685. The motifs B and C are DIVLID₇₃₇D and K₇₇₄GLPFTSKIIISTSN, respectively (Figure 5A).

In the P3 region, since the cleavage sites of 3B/3Cpro and 3Cpro/3D have already been determined (Figures 2C, 3C), the only one that remained to be analyzed was at the 3A/3B junction. Results showed that the approximate 27 kDa protein band was absent in the cell-free system, indicating that the cleavage of 3A/3B is difficult to be detected in our *in vitro* assay (Figure 4H and Supplementary Figure 4B). According to the conserved cleavage sites concluded from above experiments, we speculate that the PEPT₁₀₄₈/A sequence might be the cleavage site between 3A and 3B. This prediction is consistent with the rule that the fourth residue of VPg N-terminus is a tyrosine (Y) for linkage to the 5' end of RCDaV genome (Rothberg et al., 1978). The cleavage at the T₁₀₄₈/A dipeptide can produce a 15 kDa 3A and a 2.6 kDa 3B proteins. Consequently, we conclude that RCDaV polyproteins can produce 12 mature proteins in the order: P1 region (VP0, VP3, VP1); P2 region (X1, X2, X3, 2B, 2C); P3 region (3A, 3B, 3C, and 3D) (Figure 5A).

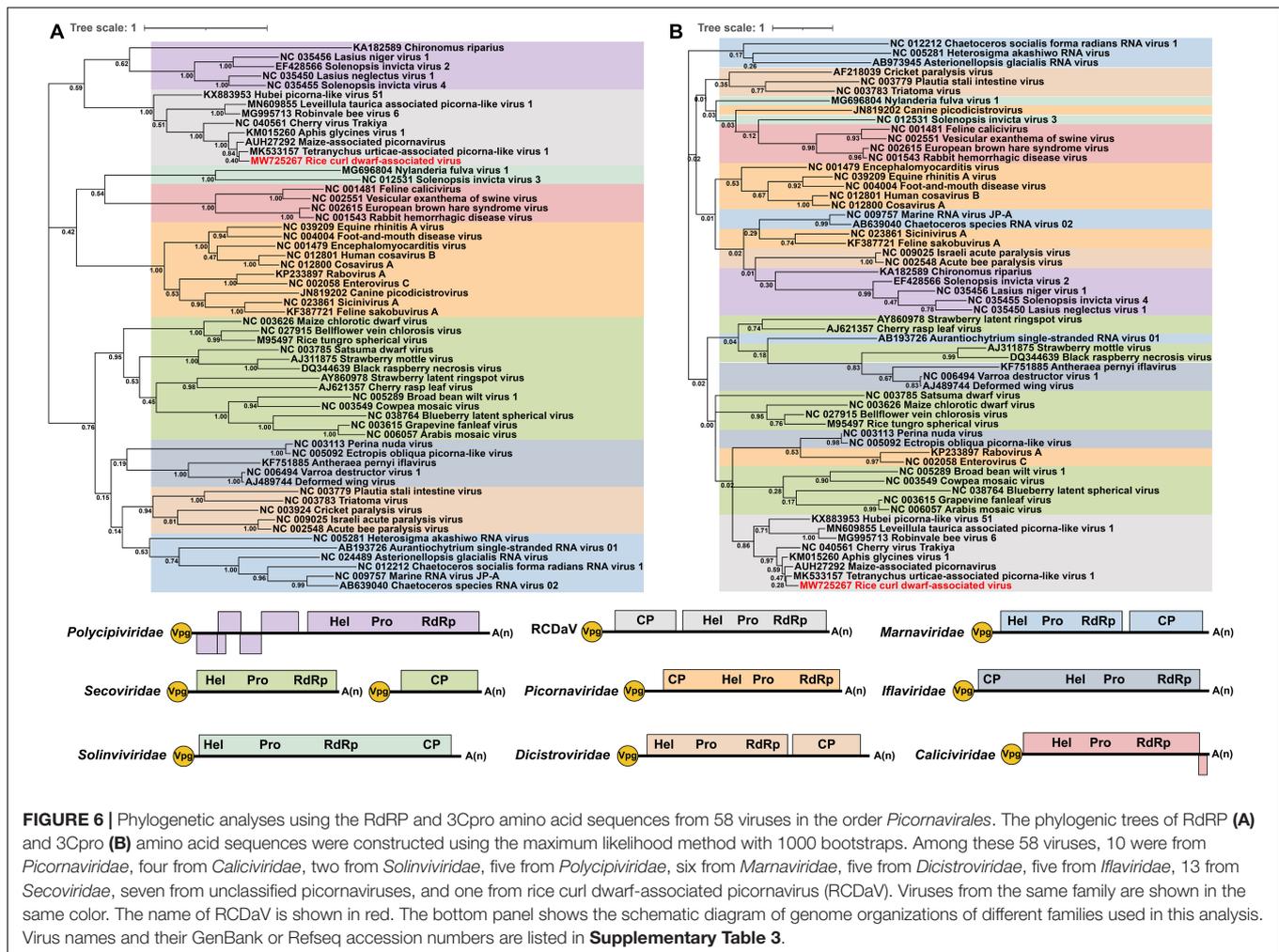
The above results of 3Cpro encouraged us to further characterize the cleavage site pattern of RCDaV. Except C/S dipeptide at the 3C/3D junction and T/A dipeptide at the 3A/3B junction, the RCDaV 3Cpro-mediated cleavage occurs at highly conserved T/S dipeptides. We performed cleavage site conservation analysis, and the result shows that the T/S dipeptide at P1/P1' position is highly conserved (Figure 5B), which differs from the classic Q(E)/G(S) cleavage sites identified in most picornaviruses. In addition, several aa residues preceding the T/S cleavage site are relatively conserved, i.e., A(V)EP (Figure 5B). To validate this finding, the cleaved products from substrate

P1-3 and P2-3 were purified and subjected to N-terminal Edman degradation sequencing. The results agreed with the finding described above, indicating the results of *trans*-cleavage assays are reliable (Figure 5C).

Phylogenetic Relationships Between RCDaV and Other Viruses in the Order Picornavirales

In addition to RCDaV, we also retrieved RdRP or 3Cpro aa sequences of 50 representative species from eight officially classified families and seven unclassified picornaviruses in the order *Picornavirales* from the NCBI database (Supplementary Table 3). Previous studies have reported that these unclassified bicistronic picornaviruses discovered from arthropods and plants that cluster into a highly divergent clade (François et al., 2019; Koloniuk et al., 2020; Yasmin et al., 2020), may represent a novel family in the order *Picornavirales* (Yasmin et al., 2020). Phylogenetic analysis indicated that RCDaV RdRP clustered together with RdRPs from these seven picornaviruses, and formed a clade distinct from eight officially classified families (Figure 6A). Moreover, the RdRP aa sequence identities between RCDaV and seven unclassified picornaviruses were 44.3–87.9%, while between RCDaV and other officially classified picornaviruses, the RdRP aa sequence identities were less than 30% (Supplementary Tables 5A,B). These results indicate that RCDaV should be grouped in this distinct clade of unclassified viruses in the order *Picornavirales*.

At the same time, the previous study observed the core motif GxSG of 3Cpros in this unclassified viruses group differs



from that of other picornaviruses (Yasmin et al., 2020). In our studies, sequence alignment results also show that RCDaV and these seven picornaviruses share the similar core motif GxSG (**Supplementary Figure 5**), suggesting that their 3Cpros are typical serine protease and different from the chymotrypsin-like cysteine protease of other picornaviruses, except 3Cpro of *Heterosigma akashiwo* RNA virus in the family *Marnaviridae*, blueberry latent spherical virus in the family *Secoviridae* and the viruses in the family *Polycipiviridae* (Olendraite et al., 2017; Mann and Sanfaçon, 2019). The phylogenetic tree based on 3Cpro aa sequences also indicates that RCDaV is phylogenetically related to the seven unclassified picornaviruses. However, 3Cpro of these eight viruses show a distant evolutionary relationship with the serine proteases of the family *Marnaviridae*, *Secoviridae*, and *Polycipiviridae* (**Figure 6B**).

This Distinct Clade of Picornaviruses Has the Conserved EPT/S Cleavage Site Pattern

Previous researchers have noticed the unusual taxonomic status of several unclassified picornaviruses and the core GxSG motif

of their 3Cpros; however, cleavage sites proteolytically processed by their 3Cpros have not been determined precisely (Yasmin et al., 2020). To find out if the highly conserved T/S dipeptide is also present in these viruses, we compared the sequences of the polyproteins of seven unclassified viruses at putative cleavage sites using cleavage positions of RCDaV polyproteins as the reference (**Figure 7**). As expected, the T/S dipeptides at the corresponding positions were highly conserved, and the aa residues flanking these cleavage sites were slightly different (**Figure 7**). We performed cleavage site conservation analysis for each junction. At some junctions, such as 3C/3D, the cleavage site contains C or T at the P1 position (**Figure 7C**), and in most cases, the 3A/3B junction contains A at the P1' position. Highly conserved EP exists at most P2 and P3 positions, with occasionally EA, EV, or VP (**Figures 7A–C**). The less conserved residues at the P4 position are V, A, or L (**Figures 7A–C**). Through alignment analysis, we found very low aa residue similarities at the P2'–P4' positions at different junctions of each virus (**Figure 7**). However, aa residues at some junctions in different viruses are similar, especially at the junctions between CPs, such as the consensus VQPT/SLIS sequence at the VP3/VP0 junction of RCDaV, MaPV, ApGIV1,

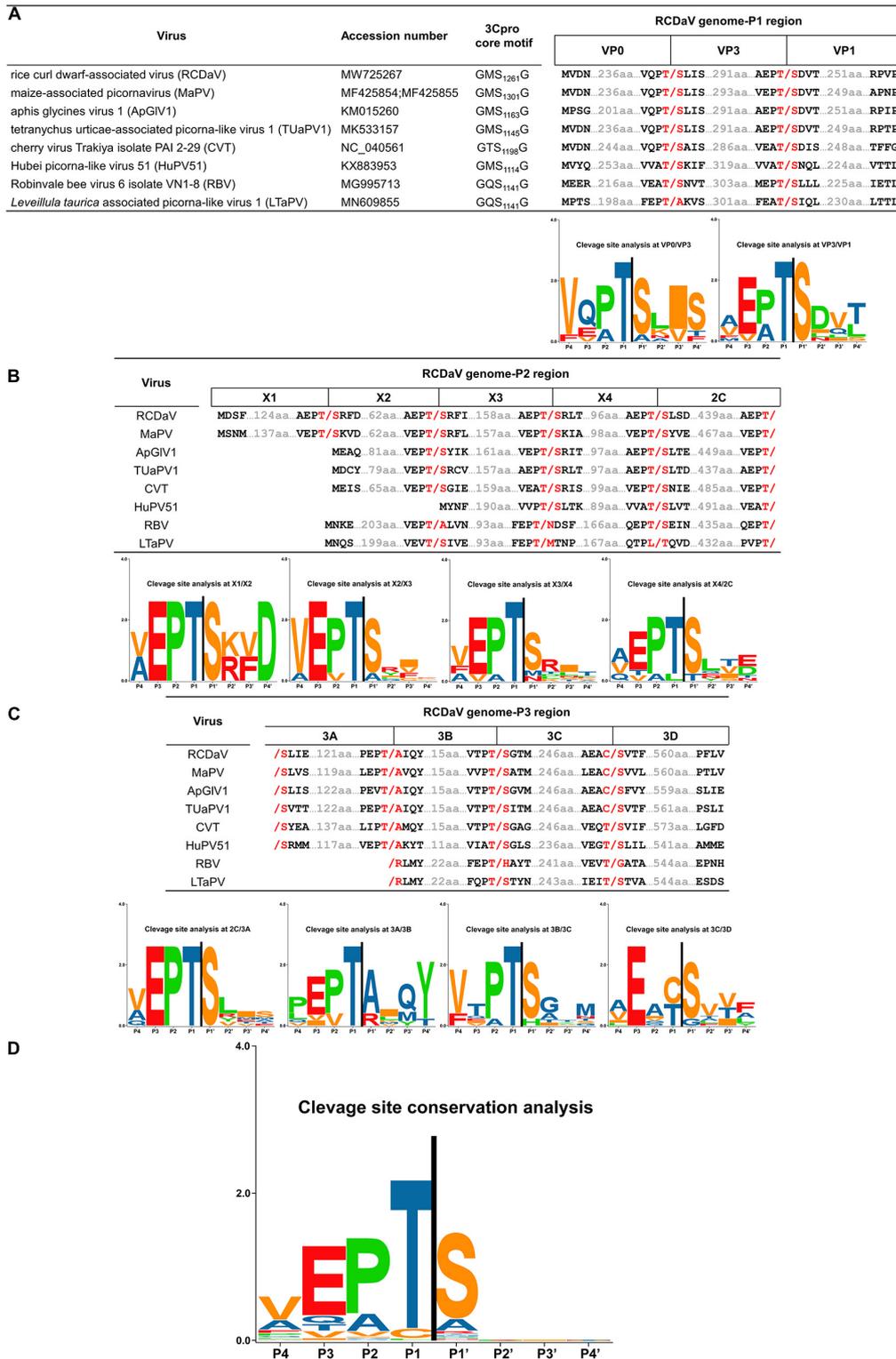


FIGURE 7 | Predicted cleavage sites in polyproteins of the seven unclassified picornaviruses based on rice curl dwarf-associated picornavirus (RCDaV) cleavage sites. **(A–C)**, Predicted cleavage sites in P1, P2, and P3 regions are shown. The GxSG motif of each virus and conservation analysis of each cleavage sites are also shown. The viral proteins were named according to the cleavage sites, the positions, and the sequence lengths of proteins. Compared with the other viruses, Robinvale bee virus 6 isolate VN1-8 (RBV) and *Leveillula taurica* associated picorna-like virus 1 (LTaPV) lack the 3A protein. **(D)** Conservation analysis of all the predicted cleavage sites in the eight picornaviruses. Amino acids are color-coded according to their physicochemical characteristics. Neutral and polar, green; basic, blue; acidic, red; neutral and hydrophobic, black. All amino acids are shown as one-letter standard code.

A	family	genus	viruses	core motif	SBP	cleavage sites
	Caliciviridae	Vesivirus	vesicular exanthema of swine virus	GDC ₁₂₂ G	GLH ₁₃₇	
	Solinviridae	Invictavirus	<i>Solenopsis invicta</i> virus 3	GDC ₁₉₀ G	GLH ₂₁₀	
	Marnaviridae	Labyrinthivirus	<i>Aurantiochytrium</i> single-stranded RNA virus 01	GLC ₁₉₂ G	GLH ₂₁₀	Q/G, Q/S, E/G, E/S...
	Dicistroviridae	Aparavirus	acute bee paralysis virus	GDC ₁₉₄ G	GIH ₂₁₃	
	Iflaviridae	Iflavirus	<i>Ectropis obliqua</i> picorna-like virus	VRC ₁₉₀ G	GFH ₂₀₄	
	Picornaviridae	Hepatovirus	hepatovirus A	GYC ₁₇₂ G	GIH ₁₉₁	
	Secoviridae	Waikavirus	rice tungro spherical virus	GFC ₂₈₅ G	GMH ₃₀₄	
			RCDaV	GMS ₁₉₀ G	GIQ ₂₀₉	
			MaPV	GMS ₁₉₀ G	GIQ ₂₀₉	
			ApGIV1	GMS ₁₉₀ G	GIQ ₂₀₉	
	unclassified	unclassified	TuaPV 1	GMS ₁₉₀ G	GIQ ₂₀₉	EPT(C)/S
			CVT	GTS ₁₉₀ G	GIQ ₂₀₉	
			HuPV51	GMS ₁₈₀ G	GIQ ₁₉₉	
			RBV	GQS ₂₀₂ G	GIL ₂₂₁	
			LTAaPV	GQS ₁₈₈ G	GIL ₂₀₇	
	Polycipiviridae	Sopolycivirus	<i>Solenopsis invicta</i> virus 2	GRS ₁₈₇ G	GIQ ₂₀₅	Not determined
		Hupolycivirus	Hubei picorna-like virus 81	GTS ₁₃₉ G	GIL ₁₅₈	Not determined
	Marnaviridae	Marnavirus	<i>Heterosigma akashiwo</i> RNA virus	GTS ₁₈₈ G	GIS ₂₀₄	Not determined
			blueberry latent spherical virus (subgroup C)	HDS ₂₀₂ G	GLH ₂₂₀	Not determined
			tomato ringspot virus (subgroup A)	EDC ₁₉₄ G	GMH ₂₁₂	Q/G, Q/S
		Nepovirus	tomato black ring virus (subgroup B)	DDC ₁₇₀ G	GML ₁₈₈	K/A(S), R/S
			grapevine fanleaf virus (subgroup C)	YDC ₁₈₀ G	AML ₁₉₈	C/A(S), R/E(G)
	Secoviridae	Satsumavirus (subgenus)	Satsuma dwarf virus	GHC ₁₈₉ G	GMC ₂₀₇	R/G, R/S, T/S

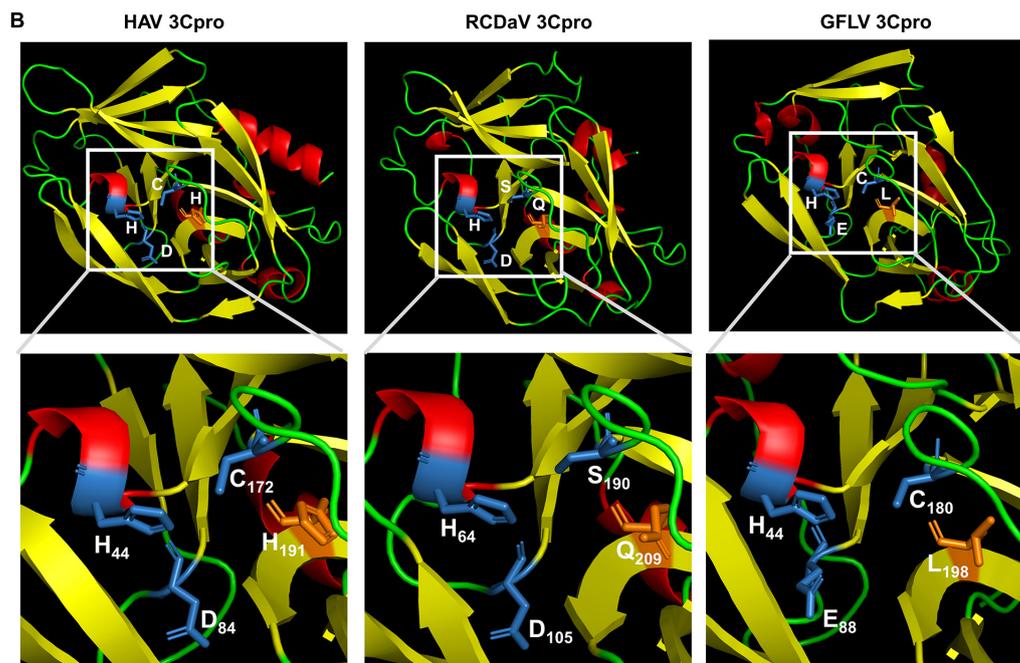
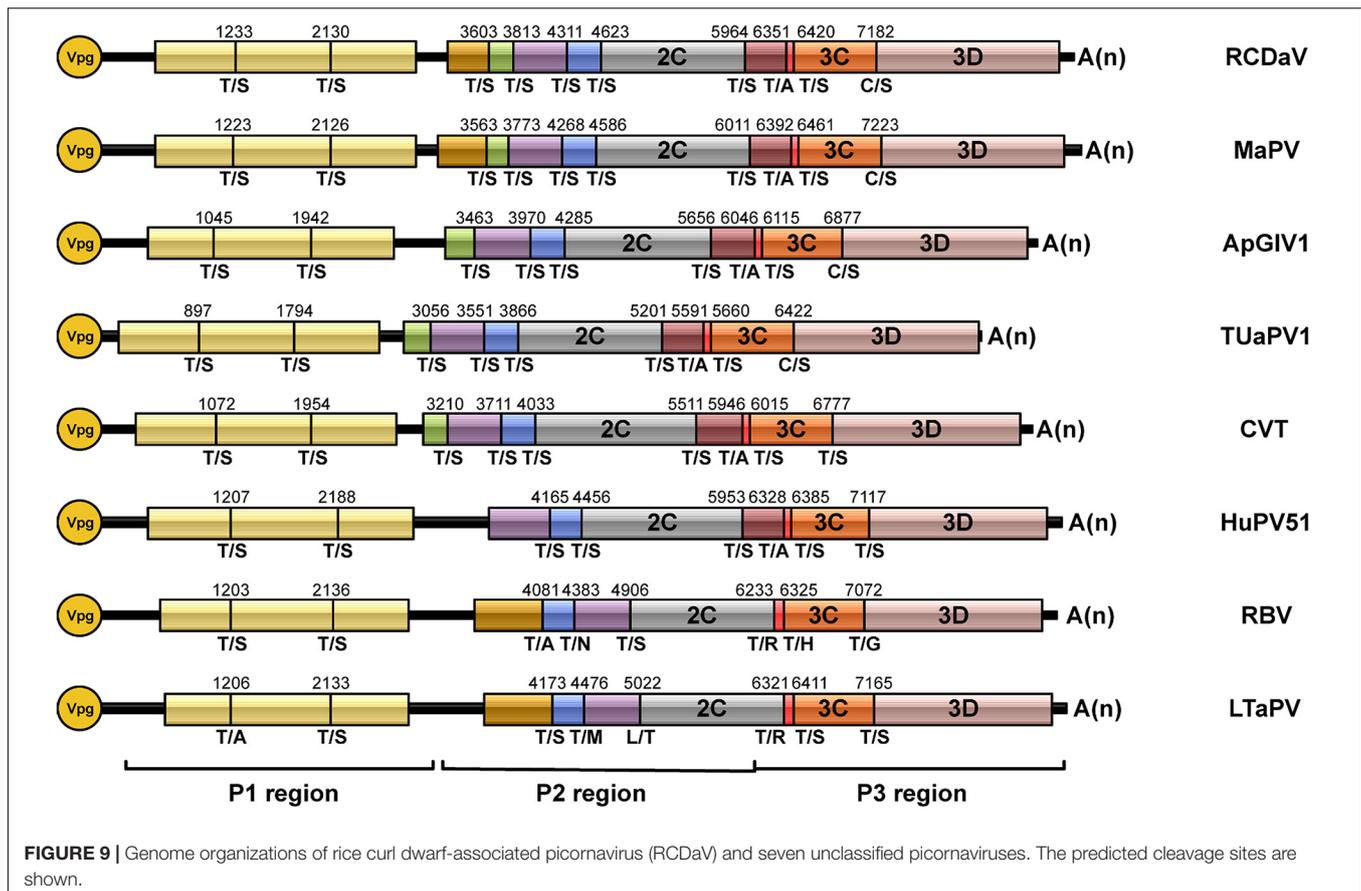


FIGURE 8 | Comparison analyses of the 3Cpro sequence and 3D structure from viruses in the order *Picornavirales*. **(A)** Comparison results of the core motif, partial substrate binding pocket (SBP) and cleavage sites of 3Cpros from viruses in the order *Picornavirales*. The cleavage sites of viruses from eight official families were obtained or predicted from previous studies (Allaire et al., 1994; Hemmer et al., 1995; Blom et al., 1996; Thole and Hull, 1998; Ye et al., 2012; Kristensen et al., 2018; Mann and Sanfaçon, 2019; Sanfaçon et al., 2020). **(B)** The 3D structure comparison of 3Cpros from HAV, RCDaV and GFLV. The 3D structure of HAV 3Cpro was downloaded from the RCSB PDB database (<https://www.rcsb.org/>, ID: 1QA7). The 3D structures of 3Cpros from RCDaV and GFLV were predicted using the HAV 3Cpro as the template through SWISS-MODEL homology-modeling server. The secondary structures are shown in different colors: red, helix; yellow, sheet; green, loop. The residues of catalytic triad are highlighted by blue and the key residues of S1 position of the SBP are highlighted in orange.

and TUAaPV1 (Figure 7A). Also the consensus sequence at their VP0/VP1 junction is A(V)EPT/SDVT (Figure 7A). Nevertheless, aa residues of CVT, Hubei picorna-like virus 51 (HuPV51), Robinvale bee virus 6 isolate VN1-8 (RBV) and *Leveillula taurica* associated picorna-like virus 1 (LtaPV) are more variable at the P2'-P4' position, probably due to the more distant homology

relationships (Figure 7). The conservation analysis for all the sites reveals that the EPT/S sequence is the conserved cleavage site of these eight viruses (Figure 7D).

Previous studies reported the conserved His in the S1 position of the substrate binding pocket (SBP) of most chymotrypsin-like viral proteases has a connection with the recognition of cleavage



sites with Gln (or Glu) at the P1 position (Mann and Sanfaçon, 2019). While the cleavage specificity of 3Cpros encoded by some nepoviruses (e.g., grapevine fanleaf virus, GFLV) may be different due to the replacement of His with Leu in the S1 position of the SBP (Margis and Pinck, 1992). We also analyzed the 3Cpro aa sequences of the eight viruses and found that the conserved His changes to Gln or Leu in the S1 position of the SBP, which could be one of the reasons why 3Cpro of the eight viruses recognize different cleavage sites (Figures 8A,B).

According to the predicted cleavage sites, we illustrated the precise genome organizations of these viruses (Figure 9). To verify these predictions for T/S cleavage sites, we analyzed two viruses (MaPV and ApGIV1). The substrate MaPV-P2₁₋₃₁₀ contains the 1–310 aa residues of MaPV P2 region, and the substrate ApGIV1-P2₁₋₁₉₉ contains the 1–199 aa residues of ApGIV1 P2 region. These two substrate genes were synthesized, and inserted in the modified pET32a vector, respectively (Figure 10A). The *trans*-cleavage assays using the cell-free system as described above showed that the 56 kDa MaPV-P2₁₋₃₁₀ protein was cleaved by MaPV 3Cpro (MBP-MaPV-Pro₁₀₉₄₋₁₄₉₀-SUMO-His) to produce 23 and 15 kDa proteins (Figure 10B). T₁₄₅S(AA) mutant abolished the production of the 23kDa protein, while T₂₁₅S(AA) mutant abolished the production of the 15 kDa protein (Figure 10B), suggesting that EPT₁₄₅/S and EPT₂₁₅/S are the cleavage sites of MaPV 3Cpro. Mutations of other predicted sites had no effect on the cleavage activity of

MaPV 3Cpro (Figure 10B). As expected, the ApGIV1-P2₁₋₁₉₉ substrate and its Q₅₂S(AA), C₅₄S(AA), and Q₁₂₈A(AA) mutants were cleaved by ApGIV1 3Cpro (MBP-ApGIV1-Pro₉₅₆₋₁₃₅₅-SUMO-His) to produce a 16 kDa protein band, while the T₈₉S(AA) mutant was not cleaved (Figure 10C), suggesting that EPT₈₉/S is the site cleaved by ApGIV1 3Cpro. These results confirm that the EPT/S sequence is the conserved cleavage site recognized by 3Cpros from viruses in this group.

DISCUSSION

Proteases are classified in clans and families, based on their catalytic types, phylogeny, and molecular structures (Rawlings et al., 2018; Mann and Sanfaçon, 2019). A recent study indicated that the plant positive-stranded RNA virus-encoded proteases can be grouped into two clans: chymotrypsin-like cysteine or serine proteases clan (clan PA) and the papain-like cysteine proteases clan (clan CA) (Mann and Sanfaçon, 2019). The structure of the cellular chymotrypsin is characterized by a double-barrel fold, which is shared by the viral chymotrypsin-like proteases (Mann and Sanfaçon, 2019). The activity of chymotrypsin depends on the catalytic triad containing His, Asp/Glu, and nucleophile Ser, which are brought together in the 3D structure (Dessens and Lomonosoff, 1991; Sárkány and Polgár, 2003). The Ser is conserved in the viral serine proteases,

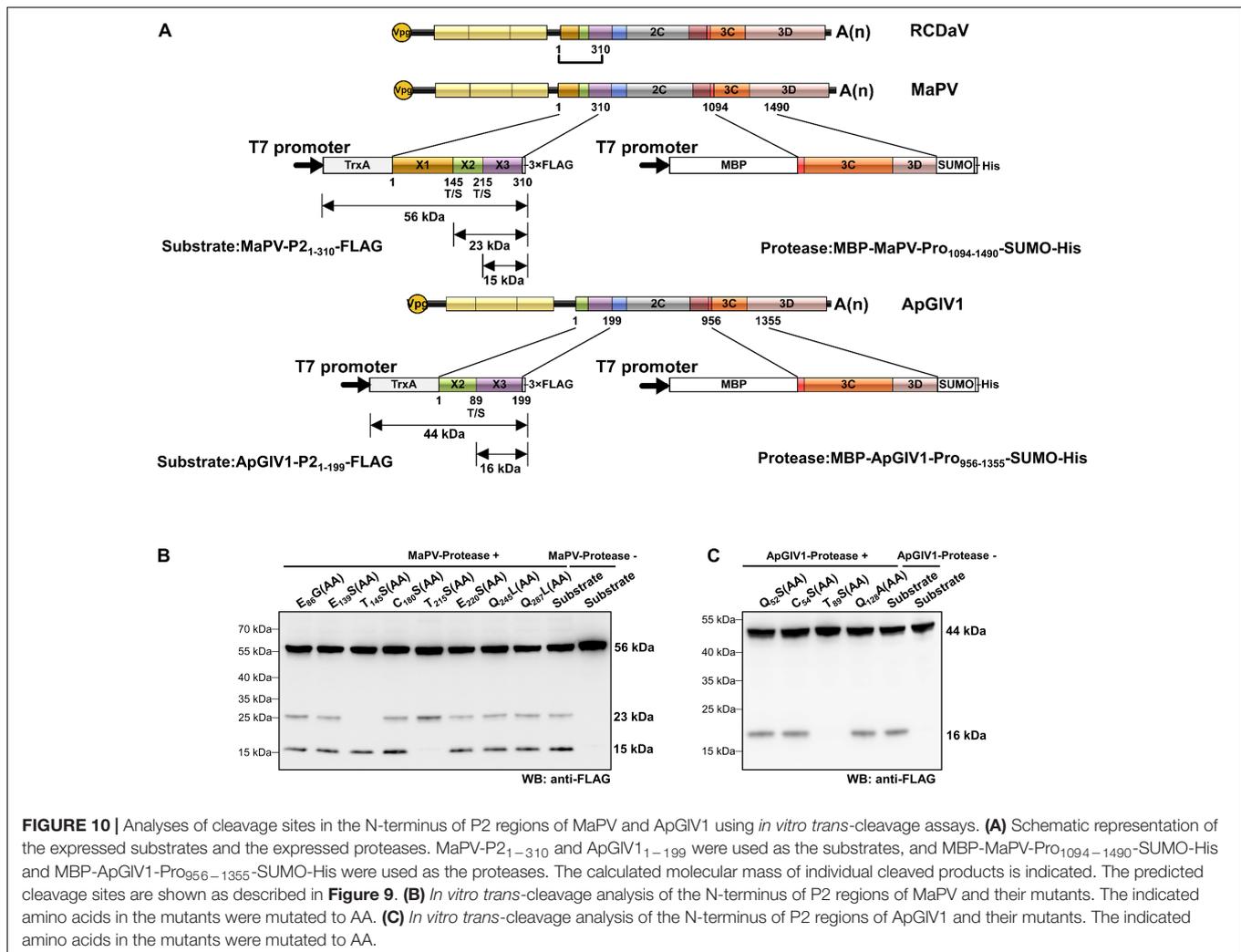


FIGURE 10 | Analyses of cleavage sites in the N-terminus of P2 regions of MaPV and ApGIV1 using *in vitro* trans-cleavage assays. **(A)** Schematic representation of the expressed substrates and the expressed proteases. MaPV-P2₁₋₃₁₀ and ApGIV1₁₋₁₉₉ were used as the substrates, and MBP-MaPV-Pro₁₀₉₄₋₁₄₉₀-SUMO-His and MBP-ApGIV1-Pro₉₅₆₋₁₃₅₅-SUMO-His were used as the proteases. The calculated molecular mass of individual cleaved products is indicated. The predicted cleavage sites are shown as described in **Figure 9**. **(B)** *In vitro* trans-cleavage analysis of the N-terminus of P2 regions of MaPV and their mutants. The indicated amino acids in the mutants were mutated to AA. **(C)** *In vitro* trans-cleavage analysis of the N-terminus of P2 regions of ApGIV1 and their mutants. The indicated amino acids in the mutants were mutated to AA.

however, in picornaviral 3Cpros, the nucleophile Ser is replaced by Cys to form the classic core motif GxCG (Allaire et al., 1994). It has been shown that this core motif is essential for the 3Cpro activity (Dessens and Lomonosoff, 1991; Blair et al., 1996; Thole and Hull, 1998). In this study, we identified a novel rice-associated picornavirus, RCDaV, which encodes a serine 3Cpro with the catalytic triad composed of His₁₁₃₅, Asp₁₁₇₆ and Ser₁₂₆₁ (**Figure 2B**). And the mutation of GMS₁₂₆₁G to GMA₁₂₆₁G abolished the *cis*- and *trans*-cleavage activities of RCDaV 3Cpro (**Figure 2B, 3B**, left panel). Meanwhile, the substitution of His₁₁₃₅ and Asp₁₁₇₆ with Aln also abolished the proteolytic activity of RCDaV 3Cpro (**Figure 2B**). Based on these findings, RCDaV 3Cpro is a chymotrypsin-like serine protease which is different from the chymotrypsin-like cysteine proteases of other picornaviruses, except 3Cpros of Heterosigma akashiwo RNA virus in the family *Marnaviridae*, blueberry latent spherical virus in the family *Secoviridae* and the viruses in the family *Polycipiviridae* (**Figure 8A**; Olenraite et al., 2017; Mann and Sanfaçon, 2019).

Cleavage sites of most picornaviral 3Cpros are known as Q(E)/G(S) dipeptides (Seipelt et al., 1999; Sanfaçon, 2015).

However, using *in vitro* cleavage assays, the conserved EPT/S cleavage sites recognized by RCDaV 3Cpro were identified (**Figure 4**), and the viral polyproteins can be processed into 12 mature proteins via *cis*- or *trans*-cleavage manner (**Figures 2-4**). Up to now, very few cleavage site proteolytically processed by serine protease in picornaviruses has been experimentally verified (**Figure 8A**). Additionally, one T/S cleavage site has been identified in satsuma dwarf virus (SDV) catalyzed by its cysteine protease previously (Iwanami et al., 1998). However, the preceding aa sequence of T/S cleavage site from SDV is AQ which has no similarity with that of RCDaV (EP), and the existence of other R/G(S) cleavage sites indicates that the T/S cleavage site is not conserved in SDV (Iwanami et al., 1998). Therefore, compared with the previous identified 3Cpro cleavage sites, the highly conserved cleavage site EPT/S of RCDaV 3Cpro is relatively uncommon in picornaviruses.

Allaire et al. (1994) have shown that the specific recognition of Gln (Q) or Glu (E) at the P1 position is conferred by the conserved His in the S1 position of the SBP of most chymotrypsin-like viral 3Cpros. Although some 3Cpros from viruses in the family *Secoviridae* substitute Leu (nepoviruses

of subgroup A and B, sequiviruses) or Cys (satsumavirus) for the conserved His in their SBPs (Mann and Sanfaçon, 2019; **Figure 8A**), and this may lead some nepovirus proteases recognize a variety of different cleavage sites with Asn, Asp, Arg, Lys, Cys, or Gly at the P1 position, and sequivirus proteases recognize Asp, Ser (Sanfaçon et al., 2009), while satsumavirus proteases recognize Thr or Arg (Sanfaçon et al., 2020). In our study, the conserved His in the S1 position of the SBP of RCDaV 3Cpro is replaced by Gln (**Figure 8A**). The comparison analysis of 3D structures of 3Cpros from HAV, RCDaV, and GFLV shows that the Gln occupies the position of the His in RCDaV 3Cpro (**Figure 8B**). Moreover, sequence alignment analyses reveal that the conserved His in this position of MaPV, ApGIV1, TUaPV1, CVT, and HuPV51 is also replaced by Gln (**Figure 8A**). To our surprise, this position of 3Cpros from RBV and LtaPV is replaced by Leu, which is similar with some viruses in the family *Secoviridae* mentioned above. However, 3Cpros from RBV and LtaPV still recognize the EPT/S cleavage sites, which differ from the 3Cpros with Leu in their S1 position of the SBP of some secoviruses. Thus, the reason underlying the recognized Thr or Cys in the P1 position may be complex and remains to be further explored.

In recent years, a great number of new viruses were discovered through RNA-seq technology. Among these new viruses, picornaviruses account for a considerable proportion. Yasmin et al. (2020) identified a novel picorna-like virus ApGIV1 through high-throughput sequencing. They have found that ApGIV1 differs from viruses in the officially classified families of picornaviruses, and is phylogenetically closely related to a clade of unclassified viruses with similar characteristics. Therefore, they considered these viruses might represent a new family under the order *Picornavirales*. Yasmin et al. (2020) also noticed that the core motif of 3Cpros from these viruses have changed to the GxSG type. However, their predicted cleavage sites on the polyproteins of ApGIV1 were not mapped correctly based on our results (**Figure 7**), probably because they did not realize that the cleavage specificity of this type of 3Cpros might be different from the classic picornaviral 3Cpros. Thus, how the polyproteins of these unclassified viruses are processed into mature proteins is still unknown, which limits the further investigation of them. In our study, we prove that RCDaV clusters together with ApGIV1 into this distinct clade (**Figure 6A**), and all the viruses in this clade encode the similar serine proteases (**Figure 6B**). More importantly, we precisely mapped the EPT/S cleavage sites on the polyproteins of these picornaviruses (**Figures 7A–C**), suggesting the cleavage specificity of this type of serine protease is highly conserved. Based on these EPT/S cleavage sites, we illustrated the precise genome organizations of these viruses (**Figure 9**), and several predicted cleavage sites of MaPV and ApGIV1 were also verified experimentally (**Figures 10A,B**). These results indicate that EPT/S is the conserved cleavage site on the polyproteins from picornaviruses in this clade. The results in this study further support the previous view that this clade of picornaviruses may represent a new family, and we move forward demonstrating that their 3C serine proteases and their cleavage sites might be an essential characteristic of this clade of picornaviruses.

In summary, we have identified a novel picornavirus, rice curl dwarf-associated virus (RCDaV). Amino acid sequence alignment of RdRPs showed that RCDaV and seven picornaviruses share relatively high sequence identities ranging from 44.3 to 87.9% (**Supplementary Table 5B**). RCDaV and seven unclassified picornaviruses cluster into an independent clade which is distinct from eight officially classified families in the order *Picornavirales*. These viruses share similar genome organizations and encode the similar functional 3Cpros which are chymotrypsin-like serine proteases recognizing the conserved EPT/S cleavage sites. However, the molecular mechanism of how these 3Cpros recognize and cleave this uncommon EPT/S sites remains unclear. The impact of RCDaV on rice production, and its genomic function as well as its transmission mode in fields require further studies.

DATA AVAILABILITY STATEMENT

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found below: <https://www.ncbi.nlm.nih.gov/genbank/>, MW725267.

AUTHOR CONTRIBUTIONS

TZ and CL: writing—original draft preparation. DW: methodology. MC: validation. QW, YX, and SG: software. SF: formal analysis. JW and XZ: writing—review and editing. JW: funding acquisition. All authors have read and agreed to the published version of the manuscript.

FUNDING

This work was supported by the National Natural Science Foundation of China (Nos. 31772125 and 31972234), the National Key Research and Development Project of China (Nos. 2016YFD0300706 and 2017YFD0201604), and the Earmarked Fund for Modern Agro-industry Technology Research System (nycytx-001).

ACKNOWLEDGMENTS

We thank Li Xie and Yaqin Wang (Zhejiang University) for technical support in transmission electron microscopy; Dr. Xinchun Ding [Noble Research Institute (Retired), Ardmore, United States], Dr. Yi Xu (Nanjing Agriculture University, China), and Dr. Tao Hu (Zhejiang University) for valuable suggestions to this manuscript.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fmicb.2021.757451/full#supplementary-material>

Supplementary Figure 1 | Rice curl dwarf-associated picornavirus (RCDAV) was detected by RT-PCR in symptomatic rice and barnyard grass plants using RCDAV specific primers.

Supplementary Figure 2 | Protein 3D structure prediction of rice curl dwarf-associated picornavirus (RCDAV) 3Cpro. **(A)** Using the structure of HAV 3Cpro as the template, the 3D structure of RCDAV 3Cpro was predicted through SWISS-MODEL homology-modeling server. The color on the sequence indicates the model quality, orange, low quality; blue, high quality. The mutated amino acid sites in *cis*-cleavage assays to verify catalytic triad of RCDAV 3Cpro are indicated by arrows. The red arrows represent the catalytic triad of RCDAV 3Cpro, while greens ones represent non-effective sites. **(B)** The comparison of the 3D structure of 3Cpros from RCDAV and HAV. The secondary structures are shown in different colors: red, helix; yellow, sheet; green, loop. The residues of catalytic triad are highlighted by blue.

Supplementary Figure 3 | *Trans*-cleavage activities of tag-free rice curl dwarf-associated picornavirus (RCDAV) 3Cpro_{1072–1325}. **(A)** Schematic representation of tag-free RCDAV 3Cpro_{1072–1325}. **(B)** Substrate Pro_{1054–1496}-FLAG(S_{1261A}) was incubated with MBP-Pro_{1054–1496}-SUMO-His or tag-free RCDAV 3Cpro_{1072–1325} in the *trans*-cleavage assay followed by western blot analysis.

Supplementary Figure 4 | The original images of **Figures 4F (A)** and **4H (B)**.

Supplementary Figure 5 | Sequence alignment of 3Cpros from rice curl dwarf-associated picornavirus (RCDAV) and the seven unclassified picornaviruses. The core motifs of these 3Cpros are highlighted by red box.

Supplementary Table 1 | All primers used in this study.

Supplementary Table 2 | Details of the expressed genes and their vectors.

Supplementary Table 3 | Viruses used in the phylogenetic analyses.

Supplementary Table 4 | The BLASTp search using the RNA_dep_RNAP domain of rice curl dwarf-associated picornavirus (RCDAV).

Supplementary Table 5 | Pairwise sequence comparison of rice curl dwarf-associated picornavirus (RCDAV) with other viruses in the order *Picornavirales*. **(A)** Pairwise amino acid sequence identities (%) of the P1, P2, P3, 3C, and 3D proteins encoded by RCDAV compared to the representative members in the eight official families. **(B)** Pairwise nucleotide and amino acid sequences identities (%) of RCDAV compared with the seven unclassified picornaviruses. %N, nucleotide sequence identities (%); %A, amino acid sequence identities (%).

REFERENCES

- Allaire, M., Chernaia, M. M., Malcolm, B. A., and James, M. N. (1994). Picornaviral 3C cysteine proteinases have a fold similar to chymotrypsin-like serine proteinases. *Nature* 369, 72–76. doi: 10.1038/369072a0
- Bazan, J. F., and Fletterick, R. J. (1988). Viral cysteine proteases are homologous to the trypsin-like family of serine proteases: structural and functional implications. *Proc. Natl. Acad. Sci.* 85, 7872–7876. doi: 10.1073/pnas.85.21.7872
- Blair, W. S., Nguyen, J. H. C., Parsley, T. B., and Semler, B. L. (1996). Mutations in the poliovirus 3CD proteinase S1-specificity pocket affect substrate recognition and RNA binding. *Virology* 218, 1–13. doi: 10.1006/viro.1996.0160
- Blom, N., Hansen, J., Blaas, D., and Brunak, S. (1996). Cleavage site analysis in picornaviral polyproteins: discovering cellular targets by neural networks. *Protein Sci.* 5, 2203–2216.
- Boros, Á., Pankovics, P., Adonyi, Á., Phan, T. G., Delwart, E., and Reuter, G. (2014). Genome characterization of a novel chicken picornavirus distantly related to the members of genus *Avihepatovirus* with a single 2A protein and a megivirus-like 3' UTR. *Infect. Genet. Evol.* 28, 333–338. doi: 10.1016/j.meegid.2014.10.025
- Brown, K., Olendraite, L., Valles, S. M., Firth, A. E., Chen, Y., Guérin, D. M. A., et al. (2019). ICTV Virus Taxonomy Profile: *Soliniviridae*. *J. Gen. Virol.* 100, 736–737. doi: 10.1099/jgv.0.001242
- Chen, C., Chen, H., Zhang, Y., Thomas, H. R., Frank, M. H., He, Y., et al. (2020). TBtools: an integrative toolkit developed for interactive analyses of big biological data. *Mol. Plant* 13, 1194–1202. doi: 10.1016/j.molp.2020.06.009
- Chen, H. H., Kong, W. P., and Roos, R. P. (1995). The leader peptide of Theiler's murine encephalomyelitis virus is a zinc-binding protein. *J. Virol.* 69, 8076–8078. doi: 10.1128/JVI.69.12.8076-8078.1995
- Dessens, J. T., and Lomonosoff, G. P. (1991). Mutational analysis of the putative catalytic triad of the cowpea mosaic virus 24K protease. *Virology* 184, 738–746. doi: 10.1016/0042-6822(91)90444-G
- Devaney, M. A., Vakharia, V. N., Lloyd, R. E., Ehrenfeld, E., and Grubman, M. J. (1988). Leader protein of foot-and-mouth disease virus is required for cleavage of the p220 component of the cap-binding protein complex. *J. Virol.* 62, 4407–4409. doi: 10.1128/JVI.62.11.4407-4409.1988
- Edgar, R. C. (2004). MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* 32, 1792–1797. doi: 10.1093/nar/gkh340
- Felsenstein, J. (1985). Confidence limits on phylogenies: an approach using the bootstrap. *Evolution* 39, 783–791. doi: 10.1111/j.1558-5646.1985.tb00420.x
- Flint, M., and Ryan, M. D. (1997). Virus-encoded proteinases of the picornavirus super-group. *J. Gen. Virol.* 78, 699–723. doi: 10.1099/0022-1317-78-4-699
- François, S., Mutuel, D., Duncan, A., Rodrigues, L., Danzelle, C., Lefevre, S., et al. (2019). A new prevalent densovirus discovered in *Acari*: insight from metagenomics in viral communities associated with two-spotted mite (*Tetranychus urticae*) populations. *Viruses* 11:233. doi: 10.3390/v11030233
- Fu, S., Xu, Y., Li, C., Li, Y., Wu, J., and Zhou, X. (2018). *Rice Stripe Virus* interferes with S-acylation of remorin and induces its autophagic degradation to facilitate virus infection. *Mol. Plant* 11, 269–287. doi: 10.1016/j.molp.2017.11.011
- Grabherr, M. G., Haas, B. J., Yassour, M., Levin, J. Z., Thompson, D. A., Amit, I., et al. (2011). Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nat. Biotechnol.* 29, 644–652. doi: 10.1038/nbt.1883
- Hales, L. M., Knowles, N. J., Reddy, P. S., Xu, L., Hay, C., and Hallenbeck, P. L. (2008). Complete genome sequence analysis of seneca valley virus-001, a novel oncolytic picornavirus. *J. Gen. Virol.* 89, 1265–1275. doi: 10.1099/vir.0.83570-0
- Hemmer, O., Greif, C., Dufourcq, P., Reinbolt, J., and Fritsch, C. (1995). Functional characterization of the proteolytic activity of the tomatoblack ring nepovirus RNA-1-encoded polyprotein. *Virology* 206, 362–371. doi: 10.1016/S0042-6822(95)80051-4
- Iwanami, T., Kondo, Y., Makita, Y., Azeyanagi, C., and Ieki, H. (1998). The nucleotide sequence of the coat protein genes of satsuma dwarf virus and navel orange infectious mottling virus. *Arch. Virol.* 143, 405–412. doi: 10.1007/s007050050297
- Jones, D. T., Taylor, W. R., and Thornton, J. M. (1992). The rapid generation of mutation data matrices from protein sequences. *Bioinformatics* 8, 275–282. doi: 10.1093/bioinformatics/8.3.275
- Kjær, J., and Belsham, G. J. (2018). Selection of functional 2A sequences within foot-and-mouth disease virus; requirements for the NPGP motif with a distinct codon bias. *RNA* 24, 12–17. doi: 10.1261/rna.063339.117
- Koloniuk, I., Příbylová, J., and Špak, J. (2020). Complete genome sequence of a mite-associated virus obtained by high-throughput sequencing analysis of an apple leaf sample. *Arch. Virol.* 165, 1501–1504. doi: 10.1007/s00705-020-04620-8
- Kozak, M. (1999). Initiation of translation in prokaryotes and eukaryotes. *Gene* 234, 187–208. doi: 10.1016/S0378-1119(99)00210-3
- Kristensen, T., Newman, J., Guan, S. H., Tuthill, T. J., and Belsham, G. J. (2018). Cleavages at the three junctions within the foot-and-mouth disease virus capsid precursor (P1–2A) by the 3C protease are mutually independent. *Virology* 522, 260–270. doi: 10.1016/j.virol.2018.07.010
- Kumar, S., Stecher, G., Li, M., Niyaz, C., and Tamura, K. (2018). MEGA X: molecular evolutionary genetics analysis across computing platforms. *Mol. Biol. Evol.* 35, 1547–1549. doi: 10.1093/molbev/msy096
- Lange, J., Groth, M., Fichtner, D., Granzow, H., Keller, B., Walther, M., et al. (2014). Virus isolate from carp: genetic characterization reveals a novel picornavirus with two aphthovirus 2A-like sequences. *J. Gen. Virol.* 95, 80–90. doi: 10.1099/vir.0.058172-0
- Le Gall, O., Christian, P., Fauquet, C. M., King, A. M. Q., Knowles, N. J., Nakashima, N., et al. (2008). *Picornavirales*, a proposed order of positive-sense single-stranded RNA viruses with a pseudo-T = 3 virion architecture. *Arch. Virol.* 153, 715–727. doi: 10.1007/s00705-008-0041-x

- Letunic, I., and Bork, P. (2019). Interactive Tree of Life (iTOL) v4: recent updates and new developments. *Nucleic Acids Res.* 47, 256–259. doi: 10.1093/nar/gkz239
- Mann, K., and Sanfaçon, H. (2019). Expanding repertoire of plant positive-strand RNA virus proteases. *Viruses* 11:66. doi: 10.3390/v11010066
- Mann, K. S., Walker, M., and Sanfaçon, H. (2017). Identification of cleavage sites recognized by the 3C-like cysteine protease within the two polyproteins of strawberry mottle virus. *Front. Microbiol.* 8:745. doi: 10.3389/fmicb.2017.00745
- Marchler-Bauer, A., Derbyshire, M. K., Gonzales, N. R., Lu, S., Chitsaz, F., Geer, L. Y., et al. (2015). CDD: NCBI's conserved domain database. *Nucleic Acids Res.* 43, D222–D226. doi: 10.1093/nar/gku1221
- Margis, R., and Pinck, L. (1992). Effects of site-directed mutagenesis on the presumed catalytic triad and substrate-binding pocket of grapevine fanleaf nepovirus 24-kDa proteinase. *Virology* 190, 884–888. doi: 10.1016/0042-6822(92)90931-E
- Massart, S., Olmos, A., Jijakli, H., and Candresse, T. (2014). Current impact and future directions of high throughput sequencing in plant virus diagnostics. *Virus Res.* 188, 90–96. doi: 10.1016/j.virusres.2014.03.029
- Matthews, D. A., Smith, W. W., Ferre, R. A., Condon, B., Budahazi, G., Sllson, W., et al. (1994). Structure of human rhinovirus 3C protease reveals a trypsin-like polypeptide fold, RNA-binding site, and means for cleaving precursor polyprotein. *Cell* 77, 761–771. doi: 10.1016/0092-8674(94)90059-0
- Olendraitė, I., Lukhovitskaya, N. I., Porter, S. D., Valles, S. M., and Firth, A. E. (2017). *Polycipiviridae*: a proposed new family of polycistronic picorna-like RNA viruses. *J. Gen. Virol.* 98, 2368–2378. doi: 10.1099/jgv.0.000902
- Rawlings, N. D., Barrett, A. J., Thomas, P. D., Huang, X., Bateman, A., and Finn, R. D. (2018). The MEROPS database of proteolytic enzymes, their substrates and inhibitors in 2017 and a comparison with peptidases in the PANTHER database. *Nucleic Acids Res.* 46, D624–D632. doi: 10.1093/nar/gkx1134
- Roberts, L. O., and Groppelli, E. (2009). An atypical IRES within the 5' UTR of a dicistrovirus genome. *Virus Res.* 139, 157–165. doi: 10.1016/j.virusres.2008.07.017
- Rothberg, P. G., Harris, T. J. R., Nomoto, A., and Wimmer, E. (1978). O4-(5'-uridylyl) tyrosine is the bond between the genome-linked protein and the RNA of poliovirus. *Proc. Natl. Acad. Sci.* 75, 4868–4872. doi: 10.1073/pnas.75.10.4868
- Rueckert, R. R., and Wimmer, E. (1984). Systematic nomenclature of picornavirus proteins. *J. Virol.* 50, 957–959. doi: 10.1128/JVI.50.3.957-959.1984
- Sanfaçon, H. (2015). "Secoviridae: a family of plant Picorna-like viruses with monopartite or bipartite genomes," *Encyclopedia of Life Sciences*, ed. John Wiley & Sons, Ltd. (Chichester: John Wiley & Sons, Ltd.), 1–14. doi: 10.1002/9780470015902.a0000764.pub3
- Sanfaçon, H., Dasgupta, I., Fuchs, M., Karasev, A. V., Petrzik, K., Thompson, J. R., et al. (2020). Proposed revision of the family *Secoviridae* taxonomy to create three subgenera, "Satsumavirus", "Stramovirus" and "Cholivirus", in the genus *Sadwavirus*. *Arch. Virol.* 165, 527–533. doi: 10.1007/s00705-019-04468-7
- Sanfaçon, H., Wellink, J., Le Gall, O., Karasev, A., van der Vlugt, R., and Wetzel, T. (2009). Secoviridae: a proposed family of plant viruses within the order *Picornavirales* that combines the families *Sequiviridae* and *Comoviridae*, the unassigned genera *Cheravirus* and *Sadwavirus*, and the proposed genus *Torradorvirus*. *Arch. Virol.* 154, 899–907. doi: 10.1007/s00705-009-0367-z
- Sárány, Z., and Polgár, L. (2003). The unusual catalytic triad of poliovirus protease 3C. *Biochemistry* 42, 516–522. doi: 10.1021/bi027004w
- Seipelt, J., Guarné, A., Bergmann, E., James, M., Sommergruber, W., Fita, I., et al. (1999). The structures of picornaviral proteinases. *Virus Res.* 62, 159–168. doi: 10.1016/S0168-1702(99)00043-X
- Shi, M., Lin, X., Tian, J., Chen, L., Chen, X., Li, C., et al. (2016). Redefining the invertebrate RNA virosphere. *Nature* 540, 539–543. doi: 10.1038/nature20167
- Thole, V., and Hull, R. (1998). *Rice tungro spherical virus* polyprotein processing: identification of a virus-encoded protease and mutational analysis of putative cleavage sites. *Virology* 247, 106–114. doi: 10.1006/viro.1998.9225
- Tseng, C., and Tsai, H. (2007). Molecular characterization of a new serotype of duck hepatitis virus. *Virus Res.* 126, 19–31. doi: 10.1016/j.virusres.2007.01.012
- Wetzel, T., Chisholm, J., Dupuis-Maguiraga, L., Bassler, A., and Sanfaçon, H. (2013). *In vitro* and *in vivo* evidence for differences in the protease activity of two arabis mosaic nepovirus isolates and their impact on the infectivity of chimeric cDNA clones. *Virology* 446, 102–111. doi: 10.1016/j.virol.2013.07.040
- Woo, P. C. Y., Lau, S. K. P., Choi, G. K. Y., Huang, Y., Teng, J. L. L., Tsoi, H. W., et al. (2012). Natural occurrence and characterization of two internal ribosome entry site elements in a novel virus, canine picodistrovirus, in the picornavirus-like superfamily. *J. Virol.* 86, 2797–2808. doi: 10.1128/JVI.05481-11
- Yang, X., Cheng, A., Wang, M., Jia, R., Sun, K., Pan, K., et al. (2017). Structures and corresponding functions of five types of picornaviral 2A proteins. *Front. Microbiol.* 8:1373. doi: 10.3389/fmicb.2017.01373
- Yasmin, T., Thekke-Veetil, T., Hobbs, H. A., Nelson, B. D., McCoppin, N. K., Lagos-Kutz, D., et al. (2020). Aphis glycines virus 1, a new bicistronic virus with two functional internal ribosome entry sites, is related to a group of unclassified viruses in the *Picornavirales*. *J. Gen. Virol.* 101, 105–111. doi: 10.1099/JGV.0.001355
- Ye, S., Xia, H., Dong, C., Cheng, Z., Xia, X., Zhang, J., et al. (2012). Identification and characterization of Iflavirus 3C-like protease processing activities. *Virology* 428, 136–145. doi: 10.1016/j.virol.2012.04.002
- Yinda, C., Zell, R., Deboutte, W., Zeller, M., Conceição-Neto, N., Heylen, E., et al. (2017). Highly diverse population of *Picornaviridae* and other members of the *Picornavirales*, in cameroonian fruit bats. *BMC Genomics* 18:249. doi: 10.1186/s12864-017-3632-7
- Zhang, T., Liang, Q., Li, C., Fu, S., Kundu, J. K., Zhou, X., et al. (2020). Transcriptome analysis of rice reveals the lncRNA–mRNA regulatory network in response to *Rice Black-Streaked Dwarf Virus* infection. *Viruses* 12:951. doi: 10.3390/v12090951

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2021 Zhang, Li, Cao, Wang, Wang, Xie, Gao, Fu, Zhou and Wu. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.