

A novel *Waikavirus* (the family *Secoviridae*) genome sequence identified in rapeseed (*Brassica napus*)

D. PARK, Y. HAHN*

Department of Life Science, Chung-Ang University, Seoul 06974, Republic of Korea

Received September 7, 2018; accepted October 23, 2018

Summary. – The genome sequence of a novel species of the genus *Waikavirus* (the family *Secoviridae*), which we named Brassica napus RNA virus 1 (BnRV1), was identified in a rapeseed (*Brassica napus*) transcriptome dataset. The BnRV1 genome was 12,293 nucleotides long followed by a poly(A) tail. Two open reading frames (ORFs), called ORF1 and ORFX, were predicted. The larger ORF, ORF1, encodes a polyprotein of 3,471 amino acids and the smaller ORF, ORFX, overlaps ORF1 and encodes an 87 aa long protein of unknown function. The BnRV1 ORF1 polyprotein was predicted to undergo proteolytic processing to yield seven mature proteins, including an RNA-dependent RNA polymerase and three distinct coat proteins. The ORF1 and ORFX proteins share sequence similarities with the respective proteins of viruses in the genus *Waikavirus*, including the bellflower vein chlorosis virus, rice tungro spherical virus, and maize chlorotic dwarf virus. A phylogenetic tree inferred from a conserved segment of the polyproteins of several *Secoviridae* viruses confirmed that BnRV1 is a novel species of the genus *Waikavirus*. The BnRV1 genome sequence identified in this study may be useful for the study of waikavirus biology and waikavirus-derived diseases.

Keywords: Brassica napus RNA virus 1; Waikavirus; Secoviridae; rapeseed

Introduction

Viruses of the genus *Waikavirus* (the family *Secoviridae* of the order *Picornavirales*) have a positive-sense single-stranded RNA genome (Sanfaçon *et al.*, 2009; Thompson *et al.*, 2017). There are three known species in the genus *Waikavirus*: rice tungro spherical virus (RTSV), maize chlorotic dwarf virus (MCDV), and bellflower vein chlorosis virus (BVCoV), which are all plant pathogens (Reddick *et al.*, 1997; Sailaja *et al.*, 2013; Seo *et al.*, 2015). RTSV and MCDV are agriculturally important plant pathogens that cause severe production losses in rice and corn, respectively (Reddick *et al.*, 1997; Sailaja *et al.*, 2013).

*Corresponding author. E-mail: hahnny@cau.ac.kr; phone: +82-2-820-5812.

Abbreviations: BnRV1 = Brassica napus RNA virus 1; BVCoV = bellflower vein chlorosis virus; CP = coat protein; MCDV = maize chlorotic dwarf virus; NCR = non-coding region; NTP = nucleoside triphosphate-binding protein; ORF = open reading frame; RdRp = RNA-dependent RNA polymerase; RTSV = rice tungro spherical virus

Waikavirus genomes are composed of a mono-segmented single-stranded RNA of about 12 kilobases (kb) with a poly(A) tail, which is encapsulated in an icosahedral particle (Sanfaçon *et al.*, 2009; Thompson *et al.*, 2017). Waikaviruses are known to have at least two well-conserved open reading frames (ORFs) (Firth and Atkins, 2008; Sanfaçon *et al.*, 2009; Thompson *et al.*, 2017). The large ORF, ORF1, encodes a polyprotein that is subsequently processed into seven mature proteins, a hypothetical protein, three distinct coat proteins, a nucleoside triphosphate-binding protein, a 3C-like proteinase, and an RNA-dependent RNA polymerase. The second smaller ORF, ORFX, overlaps ORF1 and encodes an 87 aa protein of unknown function (Firth and Atkins, 2008). Possible additional small ORFs have been proposed near the 3'-end of the genome of other waikaviruses, although their existence has not been confirmed (Shen *et al.*, 1993; Thole and Hull, 1996; Reddick *et al.*, 1997; Isogai *et al.*, 2000; Verma and Dasgupta, 2007; Sanfaçon *et al.*, 2009).

Rapeseed (*Brassica napus*), also known as oilseed rape, is a bright-yellow flowering herbaceous plant of the family *Brassicaceae* (also known as *Cruciferae*). Rapeseed has

two subgenomes, one derived from *B. rapa* and the other from *B. oleracea*, indicating that it is a hybrid species that originated from a recent hybridization event (Mason and Snowden, 2016). Rapeseed is mainly cultivated to produce edible seed oil, and rapeseed oil is a rich source of natural bioactive components, including a range of antioxidants (Szydłowska-Czerniak, 2013).

High-throughput RNA sequencing (RNA-seq) is a powerful, cost-effective tool for RNA virus discovery in plants, as well as viral diagnostics and virus evolution research (Park and Hahn, 2017; Park *et al.*, 2017, 2018; Roossinck, 2017). In this paper, a novel RNA virus genome sequence was identified in a transcriptome dataset obtained from rapeseed tissues (Liu *et al.*, 2017). Sequence comparison and phylogenetic analyses showed that the novel RNA virus is a distinctive species of the genus *Waikavirus* in the family *Secoviridae*.

Materials and Methods

Data collection and sequence assembly. RNA-seq data obtained from the young floral buds of rapeseed plants (Acc. Nos. SRR2052475, SRR2052499, SRR2052502, and SRR2052505) (Liu *et al.*, 2017) were downloaded from the Sequence Read Archive (SRA) of National Center for Biotechnology Information (NCBI). The dataset contains about 19.95 gigabases from 239.7 million reads.

Raw sequence data were trimmed to collect high-quality reads using the Sickle program (version 1.33; <https://github.com/najoshi/sickle>). High-quality reads were *de novo* assembled into contigs using the SPAdes Genome Assembler (version 3.10.1; <http://spades.bioinf.spbau.ru>) (Bankevich *et al.*, 2012). Each of the four RNA-seq runs was independently assembled.

Viral data collection. A BLAST-searchable database containing representative viral RNA-dependent RNA polymerase (RdRp) motif sequences was prepared from the Pfam database (release 31.0; <http://pfam.xfam.org>). A total of 394 viral RdRp motif sequences were obtained from 19 families (PF00602, PF00603, PF00604, PF00680, PF00946, PF00972, PF00978, PF00998, PF02123, PF03431, PF04196, PF04197, PF05788, PF05919, PF07925, PF08467, PF08716, PF08717, and PF12426). A BLASTX search was performed against the representative RdRp motif database using the assembled rapeseed transcriptome contigs as queries to identify potential virus-derived sequences.

RNA sequence analysis. The rapeseed RNA-seq reads were mapped to a viral contig using the BWA program (version 0.7.17-r1194; <https://github.com/lh3/bwa>) (Li and Durbin, 2009) to collect virus-derived reads and to examine sequencing depth. The SAMtools and BCFTools programs (version 1.6; <http://www.htslib.org>) (Li *et al.*, 2009) were used to identify single nucleotide polymorphisms (SNPs). Integrative Genomics Viewer was used to visualize sequencing coverage and depth (Robinson *et al.*, 2011).

ORF and protein characterization. The getorf program (version 6.6.0.0; <http://www.bioinformatics.nl/cgi-bin/emboss/getorf>) of the EMBOSS package was used to predict ORFs and their encoded protein sequences. Sequence similarity searches of the putative viral contig against all known nucleotide and protein sequences were conducted using the BLAST programs at the NCBI website (<https://blast.ncbi.nlm.nih.gov/Blast.cgi>). The Pfam database (release 31.0; <http://pfam.xfam.org>) was utilized to characterize the functional domains in the viral genome.

Phylogenetic analysis. Multiple sequence alignments were generated using the MUSCLE program (version 3.8.31; <https://www.drive5.com/muscle>) (Edgar, 2004). A phylogenetic tree was inferred by the neighbor-joining method using MEGA (version X; <https://www.megasoftware.net>) (Kumar *et al.*, 2018).

Results and Discussion

One of the contig sequences assembled from the rapeseed (*Brassica napus*) flower bud RNA-seq set SRR2052475 contains a putative protein coding region showing high similarity with an RdRp motif (Pfam Acc. No. PF00680) of the RTSV ORF1 polyprotein (UniProt Acc. No. Q83034). RTSV is the prototype species of the genus *Waikavirus* of the family *Secoviridae* (Shen *et al.*, 1993). The contig was thought to be a genome sequence derived from a novel RNA virus related to the genus *Waikavirus* and was named *Brassica napus* RNA virus 1 (BnRV1). The BnRV1 genome sequence with annotation information has been deposited in the NCBI nucleotide database under Acc. No. MH844554.

The BnRV1 genome contig, which was assembled from 58,986 RNA-seq reads, is 12,293 nucleotides (nt) long followed by a poly(A) tail. A single nt sequence variation (G or T) was found at position 9660, which resulted in a methionine/isoleucine amino acid (aa) polymorphism in the ORF1 polyprotein. This indicated that the assembled BnRV1 genome was derived from a highly homogeneous population and may be descendant of a single clone.

Sequence comparison and ORF prediction revealed two well-conserved ORFs in the BnRV1 genome (Fig. 1), the larger ORF1 and the overlapping smaller OREFX, which were 10,416 nt (position 631–11046) and 264 nt (position 923–1186) long, respectively. The 5' non-coding region (NCR) was 630 nt long, and the 3' NCR was 1,247 nt long and was followed by a poly(A) tail. None of the other small ORFs previously suggested in other viruses of the genus *Waikavirus* were found (Shen *et al.*, 1993; Reddick *et al.*, 1997; Sanfaçon *et al.*, 2009).

BnRV1 ORF1 encoded a polyprotein of 3,471 aa. Pfam analysis showed that the polyprotein contained five conserved domains: a waikavirus capsid protein 1 (Pfam Acc. No. PF12264; aa positions 711–907), picornavirus capsid protein (PF00073; 895–1061), RNA helicase (PF00910; 1800–1906),

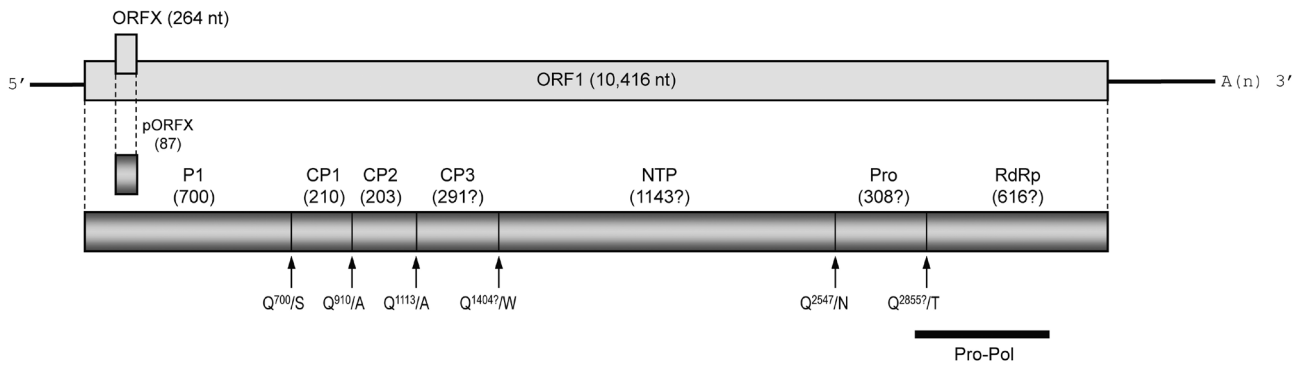


Fig. 1

Schematic representation of the BnRV1 genome

The BnRV1 genome contains two ORFs: the larger ORF1 (10,416 nt) and the overlapping smaller ORFX (264 nt), which are marked as boxes (top). ORF1 encodes a polyprotein (3,417 aa) that undergoes proteolytic cleavage to generate a hypothetical protein (P1), three coat proteins (CP1, CP2, and CP3), a nucleoside triphosphate-binding protein (NTP), a 3C-like proteinase (Pro), and an RdRp (middle). ORFX produces a small protein (pORFX, 87 aa) of unknown function. The aa residues and positions of the putative cleavage sites are marked by arrows below the polyprotein. The seven mature proteins are indicated with their respective sizes. Poorly supported sizes or positions are indicated by question marks. The “Pro-Pol” segment, which was used for the phylogenetic analysis, is marked with a line (bottom).

tungro spherical virus-type peptidase (PF12381; 2634–2864), and RdRp (PF00680; 2903–3413).

A BLASTP search was performed against the NCBI non-redundant protein database using the BnRV1 polyprotein sequence as a query. The BnRV1 polyprotein showed strong sequence similarity with the polyproteins of BVCV (Seo *et al.*, 2015), RTSV, and MCDV (Reddick *et al.*, 1997). These three viruses are species of the genus *Waikavirus*. The BnRV1 polyprotein also showed significant sequence similarity with the polyproteins of carrot necrotic dieback virus (Menzel and Vetten, 2008) and parsnip yellow fleck virus (Turnbull-Ross *et al.*, 1992). These two viruses are members of the genus *Sequivirus*, a sister taxon of the genus *Waikavirus*.

The virus with the highest similarity to BnRV1 was BVCV, and the ORF1 polyproteins showed 51% sequence identity, with almost full coverage. A BLASTN search against known viral genomes using the BnRV1 genome sequence as a query also revealed that BVCV was the closest to BnRV1, with 65% nt sequence identity and 44% coverage. Strong sequence similarity among the polyproteins of BnRV1 and other waikaviruses, including BVCV, RTSV, and MCDV, was observed in the multiple sequence alignment (Supplementary Fig. S1), indicating that BnRV1 is a novel member of the genus *Waikavirus*.

The BnRV1 polyprotein is predicted to be cleaved into seven mature proteins: a hypothetical protein (P1), three distinct coat proteins (CP1, CP2, and CP3), a nucleoside triphosphate-binding protein (NTP), a 3C-like proteinase (Pro), and an RdRp. The consensus sequence for cleavage sites is “Q/X,” where “Q” is a glutamine residue and “X” is any aa residue (Shen *et al.*, 1993; Reddick *et al.*, 1997; Sanfaçon *et al.*, 2009; Seo *et al.*, 2015). Putative cleavage sites in the

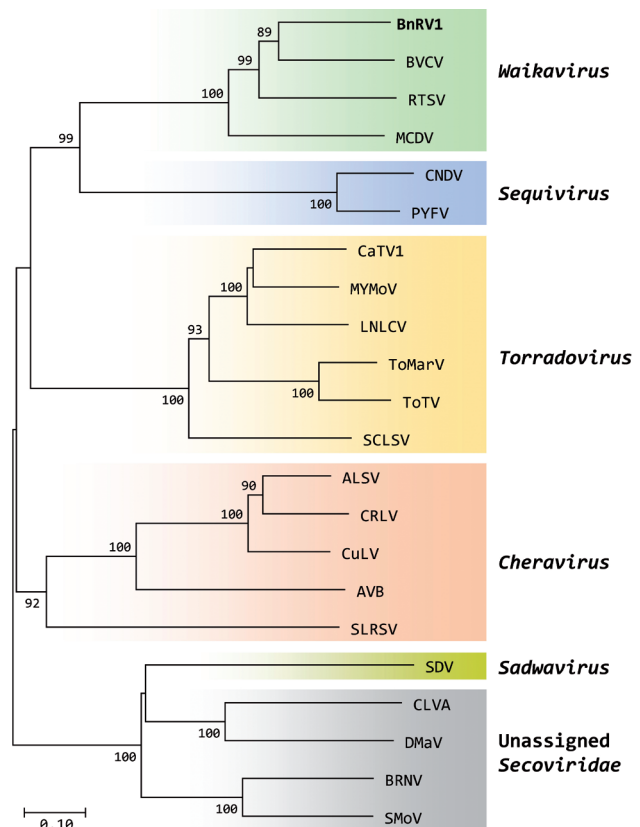


Fig. 2

Phylogenetic tree of BnRV1 and other Secoviridae viruses

A phylogenetic tree was inferred based on the “Pro-Pol” segment sequences of BnRV1 and other representatives of *Secoviridae* viruses using the neighbor-joining method. The bootstrap percentages calculated from 1,000 replicates are shown at the nodes (bootstrap support values of 80% or less are not marked). See Table 1 for the full names and Acc. Nos. of the viruses.

BnRV1 polyprotein were predicted based on comparison to the reported cleavage sites of other waikavirus polyproteins (Fig. 1). Examination of a multiple sequence alignment of waikavirus polyproteins revealed that the cleavage sites for P1/CP1, CP1/CP2, CP2/CP3, and NTP/Pro were well aligned (Supplementary Fig. S1). For these four sites, the aligned glutamine residues of the BnRV1 polyprotein were predicted to be cleavage sites. In contrast, the sites for CP3/NTP and Pro/RdRp were poorly aligned. For these two sites, nearby glutamine residues were chosen as putative cleavage sites. The six predicted cleavage sites of the BnRV1 polyprotein are as follows (the positions of the glutamine residues are given in parentheses): Q(700)/S for P1/CP1, Q(910)/A for CP1/CP2, Q(1113)/A for CP2/CP3, Q(1404)/W for CP3/NTP, Q(2547)/N for NTP/Pro, and Q(2855)/T for Pro/RdRp.

A conserved segment of the polyprotein, which is known as the “Pro-Pol” region, has been used as a taxonomical indicator for viruses of the order *Picornavirales*, including the genus *Waikavirus* (Sanfaçon *et al.*, 2009; Seo *et al.*, 2015; Thompson *et al.*, 2017). The “Pro-Pol” segment spans from the “CG” motif of the Pro to the “GDD” motif of the RdRp (aa positions 2815–3269 in the BnRV1 polyprotein). Comparison of the BnRV1 “Pro-Pol” sequence with those of representative *Secoviridae* viruses revealed a close relationship (Table 1). The BnRV1 “Pro-Pol” sequence showed 29–67% identity with those of known *Secoviridae*. Among

them, three *Waikavirus* species – BVCV, RTSV, and MCDV – showed 67%, 61%, and 57% identity, respectively. One of the criteria for species demarcation within the family *Secoviridae* by International Committee on Taxonomy of Viruses is that the “Pro-Pol” sequence shows less than 80% identity (<http://www.ictv.global/report/secoviridae>) (Thompson *et al.*, 2017). According to this criterion, BnRV1 is considered to be a novel species in the genus *Waikavirus*.

A multiple sequence alignment of the collected “Pro-Pol” sequences of representative *Secoviridae* viruses and BnRV1 was generated (Supplementary Fig. S2). A phylogenetic tree based on this alignment confirmed the classification of BnRV1 in the genus *Waikavirus* (Fig. 2).

The BnRV1 genome was predicted to have a second ORF, called ORFX, that overlaps the N-terminal region of the ORF1 polyprotein and encodes an 87 aa protein (Fig. 1). The BnRV1 ORFX protein sequence showed high sequence similarity with other waikavirus ORFX proteins (Supplementary Fig. S3). Interestingly, the sequence conservation for the ORFX proteins was higher than that for the overlapping regions of the ORF1 polyproteins. For example, the ORFX proteins of BnRV1 and BVCV showed 74% aa sequence identity, while the overlapping ORF1 polyprotein regions showed 53% identity. The ORFX proteins have only been found in the four *Waikavirus* species. No homologous proteins have been found in other viruses or other cellular

Table 1. Sequence comparison of the “Pro-Pol” segments of BnRV1 and representatives of *Secoviridae* viruses

Genus	Virus	Acronym	NCBI Acc. No.	Sequence identity ^a
<i>Waikavirus</i>	bellflower vein chlorosis virus	BVCV	NC_027915	306/455 (67%)
	rice tungro spherical virus	RTSV	NC_001632	282/463 (61%)
	maize chlorotic dwarf virus	MCDV	NC_003626	262/458 (57%)
<i>Sequivirus</i>	carrot necrotic dieback virus	CNDV	EU980442	166/487 (34%)
	parsnip yellow fleck virus	PYFV	NC_003628	161/488 (33%)
<i>Torradovirus</i>	carrot torradovirus 1	CaTV1	NC_025479	149/460 (32%)
	motherwort yellow mottle virus	MYMoV	NC_035218	142/462 (31%)
	lettuce necrotic leaf curl virus	LNLCV	NC_035214	147/462 (32%)
	tomato marchitez virus	ToMarV	NC_010987	121/385 (31%)
	tomato torrado virus	ToTV	NC_009013	120/401 (30%)
	squash chlorotic leaf spot virus	SCLSV	NC_035221	144/476 (30%)
	<i>Cheravirus</i>	apple latent spherical virus	ALSV	NC_003787
<i>Sadwavirus</i>	cherry rasp leaf virus	CRLV	NC_006271	141/457 (31%)
	currant latent virus	CuLV	NC_029038	140/457 (31%)
	Arracacha virus B	AVB	NC_020898	144/487 (30%)
	strawberry latent ringspot virus	SLRSV	NC_006964	114/386 (30%)
	satsuma dwarf virus	SDV	NC_003785	111/371 (30%)
Unassigned <i>Secoviridae</i>	chocolate lily virus A	CLVA	NC_016443	116/374 (31%)
	Dioscorea mosaic associated virus	DMaV	NC_031766	137/473 (29%)
	black raspberry necrosis virus	BRNV	NC_008182	141/470 (30%)
	strawberry mottle virus	SMoV	NC_003445	140/472 (30%)

^aAmino acid sequence identities are reported as “identical residues/aligned length (% identity).”

organisms. The strong conservation of the ORFX protein sequences suggests that they may play an important role in the biology of waikaviruses.

In conclusion, a genome sequence of a novel RNA virus, BnRV1, was identified in rapeseed. Analysis of the genomic features and phylogeny indicated that BnRV1 is a novel fourth member of the genus *Waikavirus* in the family *Secoviridae*. Three plant viruses are known to infect rapeseed – Brassica yellow virus, turnip crinkle virus, and turnip yellow virus (retrieved from the Virus-Host database at <https://www.genome.jp/virushostdb>) (Mihara *et al.*, 2016). BnRV1 is the fourth virus that is associated with rapeseed. Considering the economic value of rapeseed and the potential pathogenicity of waikaviruses, the BnRV1 genome sequence reported in this study may be useful for the study of waikavirus biology and the control of waikavirus-derived diseases.

Acknowledgment. This research was supported by the National Research Foundation of Korea funded by the Government of Korea [grant Nos. NRF-2017R1A1B4005866 and NRF-2018R1A5A1025077].

Supplementary information is available in the online version of the paper.

References

- Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, Lesin VM, Nikolenko SI, Pham S, Pribelski AD, Pyshkin AV, Sirotkin AV, Vyahhi N, Tesler G, Alekseyev MA, Pevzner PA (2012): SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. *J. Comput. Biol.* 19, 455–477. <https://doi.org/10.1089/cmb.2012.0021>
- Edgar RC (2004): MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* 32, 1792–1797. <https://doi.org/10.1093/nar/gkh340>
- Firth AE, Atkins JF (2008): Bioinformatic analysis suggests that a conserved ORF in the waikaviruses encodes an overlapping gene. *Arch. Virol.* 153, 1379–1383. <https://doi.org/10.1007/s00705-008-0119-5>
- Isogai M, Cabauatan PQ, Masuta C, Uyeda I, Azzam O (2000): Complete nucleotide sequence of the rice tungro spherical virus genome of the highly virulent strain Vt6. *Virus Genes* 20, 79–85. <https://doi.org/10.1023/A:1008116408733>
- Kumar S, Stecher G, Li M, Knyaz C, Tamura K (2018): MEGA X: Molecular Evolutionary Genetics Analysis across computing platforms. *Mol. Biol. Evol.* 35, 1547–1549. <https://doi.org/10.1093/molbev/msy096>
- Li H, Durbin R (2009): Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* 25, 1754–1760. <https://doi.org/10.1093/bioinformatics/btp324>
- Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R, 1000 Genome Project Data Processing Subgroup (2009): The Sequence Alignment/Map format and SAMtools. *Bioinformatics* 25, 2078–2079. <https://doi.org/10.1093/bioinformatics/btp352>
- Liu XQ, Yu CY, Dong JG, Xu AX, Hu SW (2017): De novo transcriptome reconstruction of a thermo-sensitive male sterility mutant in rapeseed (*Brassica napus*; Brassicaceae). *Appl. Plant Sci.* 5, apps.1700077. <https://doi.org/10.3732/apps.1700077>
- Mason AS, Snowdon RJ (2016): Oilseed rape: learning about ancient and recent polyploid evolution from a recent crop species. *Plant Biol. (Stuttg)* 18, 883–892. <https://doi.org/10.1111/plb.12462>
- Menzel W, Vetten HJ (2008): Complete nucleotide sequence of an isolate of the Anthriscus strain of parsnip yellow fleck virus. *Arch. Virol.* 153, 2173–2175. <https://doi.org/10.1007/s00705-008-0240-5>
- Mihara T, Nishimura Y, Shimizu Y, Nishiyama H, Yoshikawa G, Uehara H, Hingamp P, Goto S, Ogata H (2016): Linking virus genomes with host taxonomy. *Viruses* 8, 66. <https://doi.org/10.3390/v8030066>
- Park D, Goh CJ, Kim H, Hahn Y (2018): Identification of two novel amalgaviruses in the common eelgrass (*Zostera marina*) and in silico analysis of the amalgavirus +1 programmed ribosomal frameshifting sites. *Plant Pathol. J.* 34, 150–156.
- Park D, Hahn Y (2017): Genome sequence of spinach cryptic virus 1, a new member of the genus Alphapartitivirus (family Partitiviridae), identified in spinach. *J. Microbiol. Biotechnol.* 27, 834–837. <https://doi.org/10.4014/jmb.1611.11026>
- Park D, Kim H, Hahn Y (2017): Genome sequence of a distinct watermelon mosaic virus identified from ginseng (*Panax ginseng*) transcriptome. *Acta Virol.* 61, 479–482. https://doi.org/10.4149/av_2017_410
- Reddick BB, Habera LF, Law MD (1997): Nucleotide sequence and taxonomy of maize chlorotic dwarf virus within the family Sequiviridae. *J. Gen. Virol.* 78, 1165–1174. <https://doi.org/10.1099/0022-1317-78-5-1165>
- Robinson JT, Thorvaldsdottir H, Winckler W, Guttman M, Lander ES, Getz G, Mesirov JP (2011): Integrative genomics viewer. *Nat. Biotechnol.* 29, 24–26. <https://doi.org/10.1038/nbt.1754>
- Roossinck MJ (2017): Deep sequencing for discovery and evolutionary analysis of plant viruses. *Virus Res.* 239, 82–86. <https://doi.org/10.1016/j.virusres.2016.11.019>
- Sailaja B, Anjum N, Patil YK, Agarwal S, Malathi P, Krishnaveni D, Balachandran SM, Viraktamath BC, Mangrauthia SK (2013): The complete genome sequence of a south Indian isolate of rice tungro spherical virus reveals evidence of genetic recombination between distinct isolates. *Virus Genes* 47, 515–523. <https://doi.org/10.1007/s11262-013-0964-5>
- Sanfaçon H, Wellink J, Le Gall O, Karasev A, van der Vlugt R, Wetzel T (2009): Secoviridae: a proposed family of plant viruses within the order Picornvirales that combines the families Sequiviridae and Comoviridae, the unassigned genera Cheravirus and Sadwavirus, and the proposed genus Torradovirus. *Arch. Virol.* 154, 899–907. <https://doi.org/10.1007/s00705-009-0367-z>
- Seo JK, Kwak HR, Lee YJ, Kim J, Kim MK, Kim CS, Choi HS (2015): Complete genome sequence of bellflower vein chlorosis

- virus, a novel putative member of the genus Waikavirus. *Arch. Virol.* 160, 3139–3142. <https://doi.org/10.1007/s00705-015-2606-9>
- Shen P, Kaniewska M, Smith C, Beachy RN (1993): Nucleotide sequence and genomic organization of rice tungro spherical virus. *Virology* 193, 621–630. <https://doi.org/10.1006/viro.1993.1170>
- Szydłowska-Czerniak A (2013): Rapeseed and its products-sources of bioactive compounds: a review of their characteristics and analysis. *Crit. Rev. Food Sci. Nutr.* 53, 307–330. <https://doi.org/10.1080/10408398.2010.529959>
- Thole V, Hull R (1996): Rice tungro spherical virus: Nucleotide sequence of the 3' genomic half and studies on the two small 3' open reading frames. *Virus Genes* 13, 239–246. <https://doi.org/10.1007/BF00366984>
- Thompson JR, Dasgupta I, Fuchs M, Iwanami T, Karasev AV, Petrzik K, Sanfacon H, Tzanetakis I, van der Vlugt R, Wetzel T, Yoshikawa N, ICTV Report Consortium (2017): ICTV virus taxonomy profile: Secoviridae. *J. Gen. Virol.* 98, 529–531. <https://doi.org/10.1099/jgv.0.000779>
- Turnbull-Ross AD, Reavy B, Mayo MA, Murrant AF (1992): The nucleotide sequence of parsnip yellow fleck virus: a plant picorna-like virus. *J. Gen. Virol.* 73 (Pt 12), 3203–3211. <https://doi.org/10.1099/0022-1317-73-12-3203>
- Verma V, Dasgupta I (2007): Sequence analysis of the complete genomes of two rice tungro spherical virus isolates from India. *Arch. Virol.* 152, 645–648. <https://doi.org/10.1007/s00705-006-0861-5>

BnRV1	SSIVHRDGIAGAHLLRYSQGMPSGFSMTVIFNSFVNYYMALAWMYIVSRSELSPOADLGSFDRFTKIIVYGDDN	NVAVNDAFLDVYNLQSVACYLSLFG	3295
BVCV	SSIVHRDGLAGDVLRLYSQGMPSGFSMTVIFNSFVNYYMALSWINIANSPLSPQADLVSFYDYTKIVVYGDDN	VAVGDEFDFYNLRTVASYLSSYG	3250
RTSV	SSIVHRDGCGLDLIRYSQGMPSGFAMTVIFNSFVNYYMALAWMSTVGSLLSPQGSCKDFDYCKIVAYGDDN	VSVHEEFLDVYNLQTVAAVLSHFG	3299
MCDV	SSIIHREGIVKEYLFQYQGMPSGFSMTVIFNSFVNYYMALAMWMLNLSHSPQSTVRDFDNYCKVVVYGDDN	IVSVDLNFLEYNLRVAAYLSQFG	3267
	.*.*.*: .:.*.**:*****:.*:.*: .: * ****. .** :*:.*****:*.* ** :*.*.*.*.*.*:*		
BnRV1	ITYTDDAKNPIHLSVPHVIDSVTLKRSFVKL-DKSGSLWKAPLDKASIEERCNWIRECEPEEALNQNLLESALYEASVHGEGYFNDLKRVDGALERV		3394
BVCV	ITYTDDAKNPIHLSVPHVIDSVTLKRSFVKL-DNTGCLWKAPLDRTSIEERCNWIRECEVPQEALYQNIIESALFEASIHGKDYFMDLKNRDLFALDRV		3349
RTSV	VTYTDDAKNPIHMSKPYEDITKMSFLKRGFERV-ESSGFLWKAPLDKTSIEERLNWIRDCPTVEALEQNIIESALHEAAIHGRDYFDLVRRLNSALTRV		3398
MCDV	VTYTDDAKNPIEKSVPFVEITSVFLKRRWVPLGGRLSTIYKAPLDKTSIEERLHWIRECDNDIEALNQNIIESALYEASIHGKIYFDLQIRIACDAV		3367
	.****. **** * * . * :.**** : : . :****:***** :***:* *** **:*.*.*:.* ** * * * : * *		
BnRV1	MLPRTSEFSKSCQVRWWSNMTGAMLSQPSLASLVELSNKNHIDLGFKFNMALDGAETTLGQALGMAKNSPFIYYDV		3471
BVCV	MLPKTKESFKQCQARWWSMDMTGAILSQPALTSLEVELSKKNQIDLNLQVKDHLV-GTSMSLGDALSRAKHSPVLVFEV		3425
RTSV	MLPPTDISFEQCQARWWSVGTGA-LRAADYTSLVRRASSGHVEFNKRYRDMFR-QQDLPLKEILMKSXPVALLDLEV		3473
MCDV	MIPVPSVTFKDCCHKRWWSMTGGALDPASLRLYLAAENQLVDTRKWKDRFL-GERDSLIDMLKSARAVPLAAYHV		3443
	. . :*.*.*: ***:.*.* * . : * :.. :. : . : * : * :. . : *		

Fig. S2. Multiple sequence alignment of the “Pro-Pol” segments of BnRV1 and representative *Secoviridae* viruses

BnRV1	CG-----AAVVRADAKAIR-----KIIGMH-VAGNVSRGV--GYAEML---IKETIQTAIDALSKEVV-----QKAMDEP		
BVCV	CG-----SAIVRADTRAIR-----KVIIGMH-VAGHKQKGV--GYAETL---SLEPILEAIKRVCTPV-----ERSPVNN		
RTSV	CG-----RAIMRADATCFR-----KIIGMH-VSGLRNKCM--GYAETL---TQEHLMQAIETLKETGLLKHIPKGAIGAGEEK		
MCDV	CG-----GLILRACTRMVR-----KIVGLH-VAASANHAM--GYAEL---VQEDLKHAINKLSPDARSL-----IIGHLNP		
CNDV	CG-----SCMVSLSLEKLDG-----KVFCML-VAGTYDNL TG-QYVSTYVPVTVMNRKAISGITTLQAGD-----CQGTICD		
PYFV	CG-----SCLVSTSDKLDG-----KVFCSL-VAGTYDRVTG-KYVSTYVPITCDMIKKSISLLTGAEFSE-----SQSSICD		
CaTV1	CG-----VH-CAGYTSVMVKQGHVASY---ASVIYQEDLISMLPAVKL-----IPQAPCP		
MYMoV	CG-----IH-CAGAKAEYIRKGLKESC---ASAIYREDLEELLPKPKL-----EAQGPCA		
LNL CV	CG-----IH-CAGLKEKYVAQGSQKSY---AAAVYREDLEALLPKQML-----CSQVQCS		
ToMarV	CGT-----LLLPSVQNKQP-----VIVGIH-CASYDGAERGFISSN---ATAIYREQLDLEL-PTGPKV-----AAMVRCD		
ToTV	CGT-----LLLPLNVQDKQP-----CIVGIH-CASYDEEAHKGVFASN---ATAIFRDQLEDL-PTGPKV-----VAMVRCD		
SCLSV	CGS-----AVLIPGAINGQP-----EIVGIH-CAGFSDAMQRKGYRGT---AALIFYEDIEKYLPSVDLQ-----EAQTNP		
ALS V	CGSLGSKMPACYSYTFDTFAGLCTSP LISMDGGRVLLGLH-VAGDKSKM---GY-----AQIVTLEDFSDINFSEK-----VQGQP--		
CRLV	CGSLGSKQMPACYSYVFETYAGLCTSP LIQAEGGRICLLGLH-VVGNDRSKM---GY-----AQIVTLDLDFSDVALSDK-----VGGQP--		
CuLV	CGSLGSLRMPACYSYTFDTYPLGCTSP LICMSGGRICLLGLH-VVGNMSKT---GF-----SQIVTLDLDFSEVPLQSGTL-----VGEQP--		
AVB	CG-----DVLQVCSGVS-----KILGMHTTAGKQSIW---RY-----ANTVTLQDIEGAYPSEW-----VGEEDL		
SLRSV	CG-----TPLVANYGKGLKIASIHVVHHSATEKDTIAGSGSLLTKEEYLEASLLLDGDIKHPLETDRIQ-----ASGCLSG		
SDV	CG-----RLLCA-NLAGHW-----RVIGMC-AGEGKNRA---GVTKAL---YADIPHEFLRADNLNAVQR-----GAELDAA		
CLVA	CG-----RLLLVERHNTL-----AIAGMH-----TYGREY---PPLSGFADIPDFVTNPVI-----VAQGPSD		
DMAV	CG-----RLLALV-EKNGTL-----LIVGMH-----NFGNA---RNCAFSDIPSEFKPADELQ-----SEFSMTE		
BRNV	CG-----RLLLAHCPAGNL-----MVVGLH-VSGG-----GLNE---RATSIFFGGIDGSYKDVEDG-----KFYQQGD		
SMoV	CG-----RLLLARDEAKCL-----KIVGIH-VAGVVIQE-----KHISLFSEINGMHKTAEMA-----VQGMQEV		
	**		
BnRV1	AM-----KVCEKQCAT-----IEGKGNIGQIGMIEDKLLPKMPSRTTICKSLIHGLI-----G-DVVSEPSILSKWDRRLGEK :		
BVCV	GI-----EKCEKQCVV-----LEGKGNLGLGRVQSAVNPVPTKTTIAKSLIHGLI-----G-EVRTEPSILSKWDRKRLGDK :		
RTSV	LP-----EHSKKQSL-----LEGKGNLGLVQQLTAQLVPTSVTKTTICKSMIHGLI-----G-EIKTEPSVLSAWDRRLPFP :		
MCDV	KV-----ETATKQCGI-----VRSLSGLGCHGKVTSEDMATATKTTIRKSRIYGLV-----G-DIKTEPSILHAHDPRLPED :		
CNDV	SPVSEVVFSTLKMDELLSSTPTPSGNLGVFKPNDRMGLIEVVGRIFPNTTPKAIICRSTIIPSLIQRFM---PR----KPLTEPAILSPDLGRLGEC :		
PYFV	SPISDVAETIKVDQLFSSKPGASGKFGVGVNDTIGIIDVGRFTPETTPKSIKSTIVPSLIQPYM---PR----KPLTEPAILDPRDVLGEN :		
CaTV1	LL-----KQLRVELEN-----PFQIKQVALGQVNPDELSVSPHKTTLRQSELFHVL---SEIIG-PHLTEPSILLKGRDCKTE :		
MYMoV	QL-----RQLRQVDTN-----PFEIKQVALVGRMPPELAVNVPHKTTLRKSEIFEEL---ETHLG-PHLTEPSILTIRDTRAAS- :		
LNL CV	VL-----KKLRASTTN-----PFEIKQVAF LGTVPQELAINVPHKTTLRKSEIFDAM---TKVLG-PHSTEPSILTHHDDRPEVE :		
ToMarV	IL-----KSIRSRETQ-----LFEENQVYVLTGVPQELAAATVPHKTTLRKSQLFEAF-----G-PAETAPSILTIVHDKRGDG- :		
ToTV	LL-----KDLRARDAA-----LFEEQVAVFVGLTPAEQAATVPHKTTLRSGLF EAF-----G-PAETAPSIIASDKRGE- :		
SCLSV	AI-----EEFFQS-----PFDTQVFLGKVPRELAADIPHDALKLSIAHDIL---TEVVG-PCTTEPSILTSRDKRLSGK :		
ALS V	-----EQMYIPTAR-----SESGVSVCLLGK-WTGPKPYFLENSTLPSLIHDCI---DI----DMTTEPAILSKKDPRLQYT :		
CRLV	-----EEMYIPTKN-----SECFGSVTKLGA-WTGPKPYFLEKTSLIPSLISTSI---DV----ERTTEPAILSQDRKRLKDS :		
CuLV	-----ECYCIPTAS-----SECYGSVTKLGK-WTGPAFYFLEKTSLVPSMISKDI---DI----EMTTEPAILTTKDPRLAHS :		
AVB	PI-----SNTSR-----DFCAGEAVFGT---NEGLKPHFPPTKSLVPSIISDDLKEDDNNHTPSVMTPAILHSRDRIVEF :		
SLRSV	DE-----FFDAETVFEGLLPSQAPRQATSSIEHKSSISDGL---EKLVGKCRKTEPAIISNRDTRLKDR :		
SDV	IL-----DRFSIPISECR-----KELTPMTTRLGY-VVGQYPRALRKTIVPSIIHDNL----WR-KPETEPTILGKIDDRSPFP :		
CLVA	D-----AEFIYFDAE-----EETPMVSKIGYV---STAVPQLSKSQIEESP I FDSL---CQKLG-DPKVAPTILSKHDP RPPTP :		
DMAV	TPL-----KQITEMVSQVGF L---DAPVPLRTRTQIEKSPIFDEL---EDLNG-SALTEPTILSVTDPRPPEP :		
BRNV	LD-----VHMKELEIA-----ERVDMVDKIGRVDSSQQRPSQGTSIKSEIHDDL----WR-RAECTPTVLRSDPRPEVP :		
SMoV	DI-----LELVEP-----EVKTMVIKIGHVAQQGQFRPATKTSIVKSQIHDTL----WR-APETEPTVISPVDPRVPYA :		
	* : : * . * : : *		

BnRV1 R-GTWDPVEDAVKKGIMVVPFR--EEVKEVEEHLTRVFAKRHNSLNK-----REVNTLEIGINGIDQTPFHSPHPIEMKTSAGYPYVLRTPS 2
 BVCV R-GEWDPVMEAVKKGIVATVPFP--NEIQEVETHLCGVFQNFENSLRK-----REVNDVEVINGIDRSDFWSPHPIEMKTSAGYPYVLRKPP 2
 RTSV P-GEWDPMKDAVKKYGSYLPPFT--EEIQEVENFLIKKFRKENSRR-----RNVNSLEVINGIDGSDFWSPHPIEMKTSAGYPYVLRKPP 2
 MCDV QIGKWDVPVEAALKYGTREPFPI--EEILEVEDHLSIILKGMNTLKK-----RNVNLEVINGIDQSDYWLQIETNTSPGWPYTKRKP 2
 CNDV R---YDPMIEGKIKYQDQAKPIRG--NWRKRIIDSMREQMEDWETFMVRENYD---TLDLPMHTIINGISGIEYEEPLNMSTSEGYPILSRPS 2
 PYFV R---YDPMIDGKIKYEEQARPIKI--SWRNQIIESMAAQMQDWFETFMVREGYM---TMDLPMSSVINGIDGVEYEEPLNMSTSEGYPILNRPK 2
 CaTV1 G---FDPYVAGVQKFNETACCFDM--RVAETVMESMSDDLKSLLANIQVPGGK---PVRSEEQILNGIPGEEKYDSMDFSTSCGYPFMIM-GF 2
 MYMoV ---FDPYVKGVEKFNETAACFDN--EVAQIVMQHMKASLLANLERISVPGGK---PIVRGEHEILNGIPGEEKFYDAMDSTSCGYPFSL-EF 2
 LNL CV G---FDPYVKGVEKFNETAACFDH--ETVEVALQHMSDDLVELVKVDVPGGK---PTVRSELEILNGIPGEEFYDSMDSTACGYPYTL-EF 2
 ToMarV ---FDPYVAGVMKYNETAACGFD--DIAKLAFENLKCSLLPIMRSQKIPGGR---PCERDEDVVLNGIDGCDYDGMELSTSCGYPFNKM-GM 2
 ToTV ---FDPYVAGIQKYNETAQNFDE--DIARLAYEGLRQAILPVLHSQRVPFGK---PVTQNEDEVVLNGVDGFDYFDGMELSTSCGYPYKLN-GM 2
 SCLSV A---FDPYLAGIAKFNETAHSFNM--HVAQEAFAEYMKRRLRHLKIPVPGEK---PEVRSEMVALNGIPGEEYDPMDLSTSSGWPFNKG-ER 2
 ALSV NNKDFDPFLSGMGKYAVEAHSFD--EDELFEALDRVFSEIPMFT-----CEDLSNDQVCNGIEEDDYAEGLVMQTAEGYPFCTMRPP 2
 CRLV INPEFDVFLGEMKKYAVEAHSLD---EDLEVFEDALDRVLEIPEHA-----CEDLTNDQVCNGIEEDDYAEGLVMQTAEGYPFCTMRPP 2
 CuLV NDPDFNPKFAGMLKYAVEAHSFN---EDEFEFEDALDRVFSEIPDFN-----CQELADDEVVCNGIEEDDYAEGLVMQTAEGYPFCTMRPP 2
 AVB GHEGDFDPKDGMLKYTRAGAPFDEADEDFQDALDDVFLDIGDLRTEKEIEQNQSKAICRIINENDTLNGGIEQDFEDPIVSKTAEGYPFCCQRRP 2
 SLRSV N---LDIFKKGMMKYKAVADMSPVTDEEEKIWNVTWDSIFDLPGGIQSK-----CHLSEDENLNGVNGDNEYRGMVSTSEGWPEVLRNKN 2
 SDV ---YDPYATIGEFVQEVGPIDL--SVGSDASLVVANIIGSSWKAVGKQPC-----PTVLTWEVAINGDAIIPYERLPLSTSEGYPDSQRRNF 2
 CLVA ---YDPYAGIRKFDKEVGPFDW--SEESDLQFACAEILDWTQTKKPEKFAI---ETVCSLDVAINGVHGLPFAENFPIGTSEGFVWLDRAR 2
 DMaV ---FDPYVAGIHKFEKQAGPFD--SEESDLDLAQKNISSEWERFRPEQFDV---DTVTTLEVAIQGIDGLDYAESLPIATSEGFYIERRA 1
 BRNV ---FDPYVAGIRKFEKEVGPLDF--IDEFSTESTVLVDIAEELNNKREVGGF--ELDTVLNHAANGVEGVEYAEPLVMGTSEGYPVLRQ 2
 SMoV ---FDPYTAGIMKFEKEVGPLDF--SDPDTPESTVIEDISEELLREKSLGGF--ALDTVCSNEVAINGVDSVPAERLVMGTSEGYPVLRKA 2
 : * : : * : * : *

BnRV1 GASGKKWLF--KEVGQYPSGRARYEMEDPGLIASYDEMLGQIKRGVAP--TFITVEHPKDERRKLLKIYEVNATRTFTILPPEVNIILFRQYFGDFA 2
 BVCV GATGKKWLF--NSVGCYESGREKFEMKEEQFKTSFAMRTQILNGVVP--NIMTMECPKDERRKLGKIYDNPATRTFTILPPEINILFRMFFGDFA 2
 RTSV GAQGKKYLF--EELPYPSGRPKYAMKDPLEIENYERIKEEVTSQVGP--SMTMECLDKERRKLAKEYEKPATRTFTILSPEVNIILFRQYFGDFA 2
 MCDV GAEGKKWLF--KEVGNYPGKPILEMEDSGLIESYNKMLRDAKQGVAP--IVVTVECPKDERRKLSKIYEQPATRTFTILPPEINILFRQYFGDFA 2
 CNDV DAHGKEYLFEICEDGSRAIK-----SAKLNANYEAYGSSLSSEGEF--PLISIECPKDERRALDKIYKPKTRLFILPVEFNMHARRFLDFN 2
 PYFV DAHGKEYLFETMESGERRIK-----SAKLEAHYESYGHALQSTPEF--PLICIECPKDERRALDKIYKPKTRLFILPVEFNMHARRFLDFN 2
 CaTV1 GKNKREFLD--GEPGYLLAR-----DKPVYEEYMAAMDIAISQIVM--EMVTECAKDERLPLEKIYKPKTRLFILPFHYNMLVRKYFLDFS 2
 MYMoV GKNKRGYLD--GTPGEYVLRH-----ERPVFADFIELDSEIRQGRIT--EIVSCECAKDERLPLEKIYKPKTRLFILPFHYNMLVRKYFLDFS 2
 LNL CV GSKRGRYLD--GTPGEYVLRH-----ERPVFADFIELDSEIRQGRIT--EIVSCECAKDERLPLEKIYKPKTRLFILPFHYNMLVRKYFLDFS 2
 ToMarV GMNKRREFVQSTGEGERVELKR-----DTPVFEAWEELDVQIRKGIHV--DLVTTQCAKDERLPLEKIYKPKTRLFILPFHYNMLVRKYFLDFS 2
 ToTV GTSKREFVEPSGDGRVQLKR-----TTPIFDDWEALDVEIRKGNFV--ELVTTQCAKDERLPLEKVFGRKTRLFILPFHYNMLVRKYFLDFS 2
 SCLSV GSKRGRHV--EVEGVHLLDR-----TSEAYTAYIELLQSLADGEVP--VMVTSECAKDERLPLEKIYKPKTRLFILPFHYNMLVRQYFLDFS 2
 ALSV GVTGKTWLF--AGSPGDWHIVP-----NSLLANELNVKELNLSKNIFE--PVIGIDFPKDEKVDSSKVYIKQKTRLFILPVDYNIILVRKYFLS 2
 CRLV GASGKSWLF--AGAPGDWHIVP-----GSLLANEMHKKEVAPSRGLFE--PLIGIDFPKDEKVDSSKVYIKQKTRLFILPVDYNIILVRKYFLS 2
 CuLV GATGKSWLF--CGSPGDWHILP-----GSLLNEMNKMERNLSQGIFE--PVIGIDFPKDEKVDSSKVYIKQKTRLFILPVHYNILVRKYFLS 2
 AVB GKSQKGLL--GGIPGDWSLKK-----EGPLNDAIEHLEDNLANEVFE--PLIGIDFPKDEKVMRAKVEVVKPVRLFTILPFHYNMLVRKYFLDFS 2
 SLRSV GEAGKERFL--LGLPGCYTLNR-----NLPMYQRILDL--ALSATTIP--CIVGLDTAKDERLPLSKIYQDVKTRLFILPMEVNYLVRKYFGSFV 2
 SDV GEKGRFRF--DLKGENVRVP-----TPALMEELEVLERELQKEEVLCTINTACAKDEKTPAKKVRVQPKTRIFEIFLQINIIIRRYLFWFM 2
 CLVA NESGKERFFEEVAPGHRVP-----RGSWVEDIAEIEEAAYRGDY--LTTITCAKDEKTKLDKVFVNPKTRIFEMLNFLNLVIRKYFFFWM 2
 DMaV GDTGKERF--EINGKRIP-----KGDWVQDIDQIERAAVAGNL--EIYTMACAKDEKTKKKIYETPKTRIFEIFLQINIIIRRYLFWFM 2
 BRNV GDVGFYF--SKNLHYWEL-----NEGPAKELDELEQSVAREDFDGKIITACAKDEKTKLIKVEKPKTRIFEIFLQINIIIRRYLFWFM 2
 SMoV QDTGKFRF--DKDGRWVA-----KDEVLDDLKELKELKSDDFKGGIITACAKDEKTKMKRVREVPKTRIFEIFLQINIIIRRYLFWFM 2
 : * * : * * : * * : * * : *

Fig. S3. Multiple sequence alignments of the waikavirus ORFX proteins (a) and the overlapping region of the ORF1 polyproteins (b)

Identical and similar residues in all sequences are shaded in black and gray, respectively. Note that the ORFX proteins are more well conserved than the ORF1 polyproteins.

(a)

```

BnRV1 1  MQR-VLLIASLAINMAALFMQVLGLLLKQPIILLIVGICVIMLNIFLNVLALVTKPEEDFSQFLERASAGTPLARNAERRAPLPVRERR-- 87
BVCV 1  MQR-VLLIISLAVNMIALFLFTLGLILKLPVILLIVGVFVIMLNIFLSVLALVIKPEENGSQLERVTTGTPLARVAVPRAPLPRPVRV-- 87
RTSV 1  MQRGFYLIICLSLSMIALFVLMALVFRKPVLMITLCVIMLSIFLNVLALVVRPEENFSELVARVGTGTPLARLAERRVDLPTRGRI-- 88
MCDV 1  MQR-WLLIIGLAVNMLSCFFQMLGLVFRLLPLLLVIGLCVIMLNIFLSVLALVTRSEDTLLDLMTRMHIGTPRAQNVAPVARERTLGPRSR 89
*** ** *:.:* :.* :.*:.. *:::.. : * **.***.***.* ..* : :. : * *** * . . .

```

(b)

```

BnRV1 99  AKSAINCKLGYKY-GCALYAGIGTSSQATNYSDSWNMCDNAEHLFECFSLSDQTRGRLFSVPGESVGHHTTCAKCGASCFAAGPREAL-- 185
BVCV 99  AARVINNKFSQY-DCSVSVHPGSNPEVTSHSDCWSVCNNAQHLFECVSPSDQTRGEWFPVLGKSNWHTTCSGCCASCAFATPREGV-- 185
RTSV 99  AKGFLFDNLSIPFDDCALCPNAGTRLSQTGVSHDFVCNYVEHLFECASFRETGGKFFRACSEGWHWNATCTTCGASCRFANPRENI-- 186
MCDV 178  AEVVADNRVSEY-AKLLLSNARVGVQVTPPACDWVCNNVEHLFECFGISDAQRGHITGFNDENAYWNASCAKCGACCQGANARSAIPI 266
* . . . : : . * : . :.*:.. ***.***.***.* ..* * ..* *:::.. ***.* * . * . :

```