Novel Divavirus (the family Betaflexiviridae) and Mitovirus (the family Narnaviridae) species identified in basil (Ocimum basilicum)

C. J. GOH¹, D. PARK¹, H. KIM¹, F. SEBASTIANI², Y. HAHN^{1*}

¹Department of Life Science, Research Center for Biomolecules and Biosystems, Chung-Ang University, Seoul 06974, South Korea; ²Institute for Sustainable Plant Protection, Department of Biology, Agriculture and Food Sciences, The National Research Council of Italy, Sesto Fiorentino, Italy

Received October 3, 2017; revised December 19, 2017; accepted July 10, 2018

Summary. – Transcriptome data obtained from a plant sample often contain a large number of reads that are derived from associated RNA virus genomes that were co-isolated during RNA preparation. These virus-derived reads can be assembled into a novel plant RNA genome sequence. Here, a basil (*Ocimum basilicum*) transcriptome dataset was analyzed to identify two new RNA viruses, which were named Ocimum basilicum RNA virus 1 (ObRV1) and Ocimum basilicum RNA virus 2 (ObRV2). A phylogenetic analysis of the ObRV1 RNA-dependent RNA polymerase (RdRp) motif indicated that ObRV1 is a novel species of the genus *Divavirus* of the family *Betaflexiviridae*. ObRV1 is the fourth divavirus species to be identified. The ObRV2 RdRp motif showed sequence similarity to viruses of the genus *Mitovirus* of the family *Narnaviridae*, which infect fungal mitochondria. Although most of the known mitoviruses do not produce a functional RdRp using the plant mitochondrial genetic code, the ObRV2 encodes a full-length RdRp using both the fungal and plant mitochondrial genetic codes.

Keywords: basil; Ocimum basilicum RNA virus 1; Ocimum basilicum RNA virus 2; Divavirus; Mitovirus

Introduction

Basil (*Ocimum basilicum* L.) is a popular herb that belongs to the genus *Ocimum* of the family *Lamiaceae*. It is widely cultivated in Asia, the Mediterranean region, and the United States of America for commercial and industrial purposes (Dudai *et al.*, 2002; Suppakul *et al.*, 2003; Chiang *et al.*, 2005). Basil is also used as a medicinal, antiviral, antibacterial, and antimicrobial agent, with its essential oils being known to have many beneficial compounds, such as monoterpenoids, which act against pathogens such as virus, bacteria, and fungi (Edris and Farrag, 2003; Suppakul *et al.*, 2003; Chiang *et al.*, 2005).

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

Abbreviations: ObRV1 = Ocimum basilicum virus 1; ObRV2 = Ocimum basilicum virus 2; ORF = open reading frame; RdRp = RNA-dependent RNA polymerase; SRA = Sequence Read Archive

Transcriptome analyses of basil have been conducted to investigate its genetic features, such as the genes involved in secondary metabolism and production of unique essential oils (Rastogi *et al.*, 2014; Torre *et al.*, 2016). The plant transcriptome data often contain sequences derived from RNA viruses, which can be identified to be novel plant RNA viruses (Nouri *et al.*, 2016; Visser *et al.*, 2016; Park and Hahn, 2017a,b). In this study, a previously reported transcriptome dataset collected from red and green morphs of sweet basil was analyzed and potential sequences of two novel RNA viruses, a divavirus and a mitovirus, were identified (Torre *et al.*, 2016).

Currently, only three species are reported in the genus *Divavirus* of the family *Betaflexiviridae* (Wylie and Jones, 2011; Wylie *et al.*, 2013). The *Betaflexviridae* viruses have a linear, single-stranded, positive-sense RNA genomic segment 6.5–9.0 kb in size, which encodes 2–6 proteins, including the RNA-dependent RNA polymerase (RdRp) and several other proteins depending on the genus (Adams, 2012; Igori *et al.*, 2016).

Acronym	Full name	Acc. No.	Length (nt)	ORF	Position	Length (aa)
ObRV1	Ocimum basilicum RNA virus 1	MF196913	6930	Replicase/Coat protein	44-6721	2255
				Movement protein	5050-6219	389
ObRV2	Ocimum basilicum RNA virus 2	MF196914	2784	Replicase	347-2629	760

Table 1. Summary of RNA viruses identified in this study

Viruses of the family *Narnaviridae*, to which the genus *Mitovirus* belongs, have an unencapsidated, linear, singlestranded, positive-sense RNA genome 2.5–2.9 kb in size, which encodes a single open reading frame (ORF) for the replicase containing the RdRp motif (Hillman and Cai, 2013). Mitoviruses replicate in the cytosol or mitochondria of infected fungi and have been mainly isolated from plant-pathogenic fungi (Ghabrial and Suzuki, 2009; Xie and Ghabrial, 2012).

Materials and Methods

The previously reported basil transcriptome dataset (a total of 13 gigabase pairs), which was isolated from red and green sweet basils, was analyzed in this study (Torre *et al.*, 2016). The transcriptome data is available in the Sequence Read Archive (SRA) of the National Center for Biotechnology Information (NCBI) under the accession number SRA313233. Sickle program (version 1.33; https://github.com/najoshi/sickle) was used to collect high-quality reads with the option "-q 30 -l 55." The SPAdes Genome Assembler (version 3.10.1) was used to assemble the RNA-seq reads into contigs (Bankevich *et al.*, 2012).

The RdRp motif sequences of known RNA viruses were retrieved from the Pfam database (release 30.0; http://pfam.xfam.org). A total of 345 non-redundant RdRp motif sequences were collected from 19 RdRp families: their accession numbers are PF00602, PF00603, PF00604, PF00680, PF00946, PF00972, PF00978, PF00998, PF02123, PF03431, PF04196, PF04197, PF05788, PF05919, PF07925, PF08467, PF08716, PF08717, and PF12426.

The assembled basil transcript contigs were compared with the Pfam-derived RdRp sequences, using BLAST (Altschul *et al.*, 1990) with the parameter "-evalue e-5." Matched contigs 1000 bp or longer were selected. Annotation of contigs and identification of closely related viruses were carried out by BLAST searches of all viral genomes and proteins at the NCBI website (https://www.ncbi. nlm.nih.gov/BLAST). The basil RNA-seq reads were mapped to the viral contigs, using BWA, and variants were called using Samtools/Bcftools (Li and Durbin, 2009; Li, 2011).

For phylogenetic analysis, the RdRp motif-containing sequences were multiply aligned using the MUSCLE program (Edgar, 2004). The aligned region spanning the core RdRp motif defined by a Pfam analysis was extracted and subjected to phylogenetic analysis. A phylogenetic tree was constructed by the neighbor-joining method implemented in the ClustalW2 software (version 2.1) (Larkin *et al.*, 2007).

Results and Discussion

Genome sequences of two new viruses (Table 1) were identified by analysis of the basil transcriptome data (SRA accession number SRA313233) (Torre *et al.*, 2016). One viral contig, named Ocimum basilicum RNA virus 1 (ObRV1), was most similar to those of divaviruses (Table 2) and considered a novel member of the genus *Divavirus*. The other contig, named Ocimum basilicum RNA virus 2 (ObRV2), showed sequence similarity to those of mitoviruses (Table 2), suggesting that it belongs to the genus *Mitovirus*.

Ocimum basilicum RNA virus 1 (ObRV1)

The first of the two viral contigs discovered in the basil transcriptome was named Ocimum basilicum RNA virus 1 (ObRV1) and showed sequence similarity to those of plant RNA viruses of the genus *Divavirus* of the family *Betaflexiviridae*. There were sequence variations at 33 positions when the basil RNA-seq reads mapped to the ObRV1 genome were analyzed, indicating that the contig is a composite sequence derived from at least two closely related viruses (Supplementary Table S1).

The ObRV1 RNA genome is 6,930 base pairs (bp) long and predicted to have two overlapping ORFs. ORF1 encodes a 2,255-amino acid (aa)-long RdRp motif-containing protein and ORF2 a 389-aa-long movement protein (Table 1). A domain analysis using Pfam showed that the ObRV1 ORF1 protein contains four known domains: viral methyltransferase (PF01660) at aa 43–357, viral RNA helicase (PF01443) at aa 858–1,119, RdRp (PF00978) at aa 1,255–1,653, and trichovirus coat protein (PF05892) at aa 2,039–2,223. These genome and domain organizations are the shared features of viruses in the genera *Divavirus* and *Capillovirus* (Jelkmann, 1995; Wylie and Jones, 2011).

The ObRV1 RdRp showed 30–60% aa sequence identities with RdRps of the viruses that belong to the family *Betaflexiviridae* (Table 2). Among them, Hardenbergia virus A (HarVA), Diuris virus A (DiVA), and Diuris virus B (DiVB) were the top three similar viruses with aa sequence identities of 50–60%. These three viruses are species in the genus *Divavirus*, suggesting that ObRV1 is a new divavirus.

A phylogenetic analysis using multiply aligned RdRp motif sequences confirmed that ObRV1 is a new species of the genus *Divavirus* (Supplementary Fig. S1 and Fig. 1a).

Table 2. Identities among the RdRp sequences of ObRV1, ObRV2, and their respective related viruses

	Acronym	Full name	Acc. No. ^a	Sequence identity ^b
ObRV1	HarVA	Hardenbergia virus A	YP_004376201.1	1341/2237 (60%)
	DiVA	Diuris virus A	YP_006905850.1	1243/2248 (55%)
	DiVB	Diuris virus B	YP_006905848.1	1196/2256 (53%)
	CVA	Cherry virus A	ANE06570.1	357/949 (38%)
	CVA	Cherry virus A	APG53775.1	331/848 (39%)
	PCMV	Peach chlorotic mottle virus	YP_001497153.1	336/928 (36%)
	ApLV	Apricot latent virus	ADT91605.1	334/910 (37%)
	NeLV	Nerine latent virus	YP_009174681.1	335/1030 (33%)
	NSV	Narcissus symptomless virus	YP_842438.1	336/1030 (33%)
	CTLaV	Cherry twisted leaf associated virus	AHJ80270.1	326/840 (39%)
	CTLaV	Cherry twisted leaf associated virus	AHJ80314.1	326/859 (38%)
	CNRMV	Cherry necrotic rusty mottle virus	ALP45953.1	338/909 (37%)
	CRMaV	Cherry rusty mottle associated virus	AHA59466.2	323/846 (38%)
	CRMaV	Cherry rusty mottle associated virus	AHJ80324.1	317/839 (38%)
	CRMaV	Cherry rusty mottle associated virus	AKN20442.1	347/971 (36%)
	CGRMV	Cherry green ring mottle virus	AFU54620.1	318/825 (39%)
	BanMMV	Banana mild mosaic virus	NP_112029.1	321/870 (37%)
ObRV2	BcMV1	Botrytis cinerea mitovirus 1	CEZ26300.1	165/506 (33%)
	FpMV4	Fusarium poae mitovirus 4	YP_009272901.1	104/224 (46%)
	OMV3a	Ophiostoma mitovirus 3a	NP_660176.1	101/241 (42%)
	SsMV3	Sclerotinia sclerotiorum mitovirus 3	YP_009182164.1	110/283 (39%)
	MpMV3	Macrophomina phaseolina mitovirus 3	AMM45292.1	100/273 (37%)
	RsMV-17	Rhizoctonia solani mitovirus 17	ANR02693.1	91/230 (40%)
	RsMV-15	Rhizoctonia solani mitovirus 15	ALD89120.1	100/276 (36%)
	RsMV-14	Rhizoctonia solani mitovirus 14	ALD89119.1	102/297 (34%)
	HetMV1	Heterobasidion mitovirus 1	AIF33766.2	158/547 (29%)
	COV	Clitocybe odora virus	YP_005352912.1	174/577 (30%)
	BcMV3	Botrytis cinerea mitovirus 3	YP_009182161.1	113/304 (37%)
	FpMV3	Fusarium poae mitovirus 3	YP_009272900.1	87/219 (40%)
	SlaMV5	Soybean leaf-associated mitovirus 5	ALM62240.1	87/233 (37%)
	RsMV-19	Rhizoctonia solani mitovirus 19	ANR02690.1	87/221 (39%)
	BsMV1	Buergenerula spartinae mitovirus 1	AHY03257.1	91/242 (38%)
	CpMV1	Cryphonectria parasitica mitovirus 1-NB631	NP_660174.1	79/216 (37%)
	RcMV	Rhizoctonia cerealis mitovirus	AIT71973.1	89/241 (37%)
	SsMV15	Sclerotinia sclerotiorum mitovirus 15	AHF48631.1	81/208 (39%)
	HfMV1	Hymenoscyphus fraxineus mitovirus 1	AIU44705.1	81/208 (39%)
	SsMV16	Sclerotinia sclerotiorum mitovirus 16	AHF48632.1	89/240 (37%)
	SsMV2	Sclerotinia sclerotiorum mitovirus 2	AHX84129.1	88/232 (38%)
	SlaMV3	Soybean leaf-associated mitovirus 3	ALM62243.1	76/203 (37%)
	OMV5	Ophiostoma mitovirus 5	NP_660180.1	82/213 (38%)
	SlaMV2	Soybean leaf-associated mitovirus 2	ALM62242.1	78/209 (37%)
	FgMV1	Fusarium globosum mitovirus 1	YP_009126872.1	110/367 (30%)
	TbMV	Thielaviopsis basicola mitovirus	AAZ99833.1	127/435 (29%)

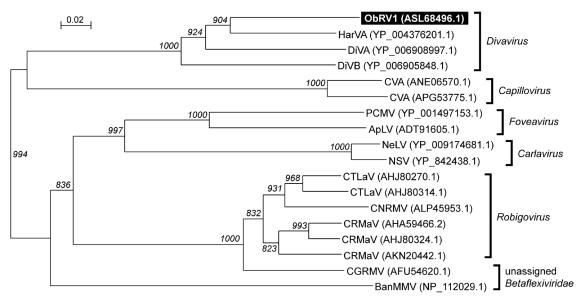
^aAcc. No. for the RdRp motif-containing protein sequence. ^bAmino acid sequence identities in a format of "identical residues/aligned length (% identity)."

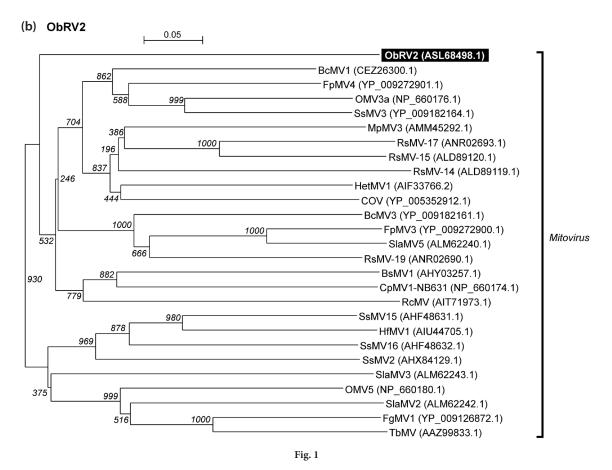
The ObRV1 was clustered with the three known divaviruses, with HarVA being the closest species. Currently, there are three known species of the genus *Divavirus* (Wylie and Jones, 2011; Wylie *et al.*, 2013); ObRV1 is the fourth divavirus species to be reported.

Ocimum basilicum RNA virus 2 (ObRV2)

The second viral contig was named Ocimum basilicum RNA virus 2 (ObRV2) and showed sequence similarity to those of viruses of the genus *Mitovirus* of the family *Nar*-







Phylogenetic analysis of newly identified basil RNA viruses

Phylogenetic trees constructed based on the amino acid sequences of the RdRp motifs of (a) ObRV1 and (b) ObRV2 are presented. The bootstrap percentages calculated from 1,000 bootstrap replicates are shown at the nodes. See Table 2 for the full names of the viruses and Supplementary Figs. S1 (ObRV1) and S2 (ObRV2) for the multiple sequence alignments of the amino acid sequences of the RdRp motifs.

naviridae. Given that mitoviruses generally infect fungi, it was uncertain whether this mitovirus genome was isolated directly from basil cells or fungal cells that infected basil (Hillman and Cai, 2013).

The ObRV2 genome is 2,784-bp long and predicted to have a single ORF encoding a 760-aa-long protein when it was translated using both the fungal mitochondrial and plant (standard) genetic codes (Table 1). A Pfam analysis indicated that the ORF contains a mitovirus-type RdRp motif (PF05919) at aa 268–579. There were sequence variations at 61 positions, indicating that the ObRV2 genome contig is also a composite sequence derived from two or more clones (Supplementary Table S2).

The ObRV2 RdRp motif-containing protein showed about 30–40% similarity to those of other mitoviruses (Table 2). A phylogenetic analysis based on a multiple sequence alignment of the RdRp domain sequences placed ObRV2 as a distinct clade within the genus *Mitovirus* (Fig. 1b and Supplementary Fig. S2).

Most of the fungal mitoviruses replicate within the mitochondria of fungal cells and are unable to produce functional RdRps in plant cells because the genetic codes of fungal mitochondria and plants are different. For instance, the UGA codon, which is a stop codon in the standard genetic code, encodes for tryptophan in fungal mitochondria (Nibert, 2017). However, the ObRV2 ORF encodes a full-length RdRp using the plant nuclear and mitochondrial genetic codes, suggesting that ObRV2 may be able to replicate in both fungal and plant cells.

Interestingly, there are mitovirus-derived sequences in many plant nuclear and mitochondrial genomes, some of which are expressed as plant genes (Bruenn *et al.*, 2015; Xu *et al.*, 2015). Because the ObRV2 sequence was found in a plant transcriptome and encodes a full-length protein, it could also be derived from a mitovirus-like gene integrated in the basil nuclear or mitochondrial genome. However, most plant mitovirus-like genes are derived from the RdRp domain and are relatively shorter than those encoding the mitovirus RdRps; plant mitovirus-like genes encode for proteins about 500-aa long or shorter, while the mitovirus RdRps are 700-aa long or longer. The ObRV2 RdRp is 760-aa long. Therefore, it is highly likely that the ObRV2 sequence represents a mitovirus that infects the basil or a basil-associated fungus.

In conclusion, basil transcriptome data were analyzed to identify viral genome sequences representing two novel viral species of the genera *Divavirus* and *Mitovirus*. The method used in this study can be applied to other plant transcriptome data to discover plant viral genome sequences co-isolated with their host RNAs.

Acknowledgment. This research was supported by the National Research Foundation of Korea funded by the Korean Government (2017R1A1B4005866).

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

References

- Adams MJ, Candresse T, Hammond J, Kreuze JF, Martelli GP, Namba S, Pearson MN, Ryu KH, Saldarelli P, Yoshikawa N (2012): Family Betaflexiviridae. In King AMQ, Adams MJ, Carstens EB, Lefkowitz EJ (Eds): Virus Taxonomy. Classification and nomenclature of viruses. Ninth report of the International Committee on Taxonomy of Viruses. Elsevier Academic Press, San Diego, CA, pp. 920–941.
- Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990): Basic local alignment search tool. J. Mol. Biol. 215, 403–410. https://doi.org/10.1016/S0022-2836(05)80360-2
- Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, Lesin VM, Nikolenko SI, Pham S, Prjibelski 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
- Bruenn JA, Warner BE, Yerramsetty P (2015): Widespread mitovirus sequences in plant genomes. PeerJ 3, e876. https://doi.org/10.7717/peerj.876
- Chiang LC, Ng LT, Cheng PW, Chiang W, Lin CC (2005): Antiviral activities of extracts and selected pure constituents of Ocimum basilicum. Clin. Exp. Pharmacol. Physiol. 32, 811–816. https://doi.org/10.1111/j.1440-1681.2005.04270.x
- Dudai N, Chaimovitsh D, Reuveni R, Ravid U, Larkov O, Putievsky E (2002): Breeding of sweet basil (Ocimum basilicum) resistant to Fusarium wilt caused by Fusarium oxysporum f.sp. basilicum. J. Herbs Spices Med. Plants 9, 45–51. https://doi.org/10.1300/J044v09n02_07
- 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
- Edris AE, Farrag ES (2003): Antifungal activity of peppermint and sweet basil essential oils and their major aroma constituents on some plant pathogenic fungi from the vapor phase. Nahrung 47, 117–121. https://doi.org/10.1002/food.200390021
- Ghabrial SA, Suzuki N (2009): Viruses of plant pathogenic fungi. Annu. Rev. Phytopathol. 47, 353–384. https://doi.org/10.1146/annurev-phyto-080508-081932
- Hillman BI, Cai G (2013): The family Narnaviridae: simplest of RNA viruses. Adv. Virus Res. 86, 149–176. https://doi.org/10.1016/B978-0-12-394315-6.00006-4
- Igori D, Lim S, Zhao F, Baek D, Park JM, Cho HS, Kim HS, Kwon SY, Moon JS (2016): The complete sequence and genome organization of Ligustrum virus A, a novel carlavirus. Arch. Virol. 161, 3593–3596. https://doi.org/10.1007/s00705-016-3054-x
- Jelkmann W (1995): Cherry virus A: cDNA cloning of dsRNA, nucleotide sequence analysis and serology reveal a new plant capillovirus in sweet cherry. J. Gen. Virol. 76 (Pt 8), 2015–2024. https://doi.org/10.1099/0022-1317-76-8-2015

- Larkin MA, Blackshields G, Brown NP, Chenna R, McGettigan PA, McWilliam H, Valentin F, Wallace IM, Wilm A, Lopez R, Thompson JD, Gibson TJ, Higgins DG (2007): Clustal W and Clustal X version 2.0. Bioinformatics 23, 2947–2948. https://doi.org/10.1093/bioinformatics/btm404
- Li H (2011): A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. Bioinformatics 27, 2987–2993. https://doi.org/10.1093/bioinformatics/btr509
- 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
- Nibert ML (2017): Mitovirus UGA(Trp) codon usage parallels that of host mitochondria. Virology 507, 96–100. https://doi.org/10.1016/j.virol.2017.04.010
- Nouri S, Salem N, Falk BW (2016): Complete genome sequence of Diaphorina citri-associated C virus, a novel putative RNA virus of the Asian citrus psyllid, Diaphorina citri. Genome Announc. 4, e00639–00616. https://doi.org/10.1128/genomeA.00639-16
- Park D, Hahn Y (2017a): 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, Hahn Y (2017b): Genome sequences of Spinach deltapartitivirus 1, Spinach amalgavirus 1, and Spinach latent virus identified in spinach transcriptome. J. Microbiol. Biotechnol. 27, 1324–1330. https://doi.org/10.4014/jmb.1703.03043
- Rastogi S, Meena S, Bhattacharya A, Ghosh S, Shukla RK, Sangwan NS, Lal RK, Gupta MM, Lavania UC, Gupta V, Nagegowda DA, Shasany AK (2014): De novo sequencing and com-

- parative analysis of holy and sweet basil transcriptomes. BMC Genomics 15, 588. https://doi.org/10.1186/1471-2164-15-588
- Suppakul P, Miltz J, Sonneveld K, Bigger SW (2003): Antimicrobial properties of basil and its possible application in food packaging. J. Agric. Food Chem. 51, 3197–3207. https://doi.org/10.1021/jf021038t
- Torre S, Tattini M, Brunetti C, Guidi L, Gori A, Marzano C, Landi M, Sebastiani F (2016): De novo assembly and comparative transcriptome analyses of red and green morphs of sweet basil grown in full sunlight. PLOS ONE 11, e0160370. https://doi.org/10.1371/journal.pone.0160370
- Visser M, Bester R, Burger JT, Maree HJ (2016): Next-generation sequencing for virus detection: covering all the bases. Virol. J. 13, 85. https://doi.org/10.1186/s12985-016-0539-x
- Wylie S, Jones M (2011): Hardenbergia virus A, a novel member of the family Betaflexiviridae from a wild legume in Southwest Australia. Arch. Virol. 156, 1245–1250. https://doi.org/10.1007/s00705-011-0963-6
- Wylie SJ, Li H, Dixon KW, Richards H, Jones MG (2013): Exotic and indigenous viruses infect wild populations and captive collections of temperate terrestrial orchids (Diuris species) in Australia. Virus Res. 171, 22–32. https://doi.org/10.1016/j.virusres.2012.10.003
- Xie J, Ghabrial SA (2012): Molecular characterizations of two mitoviruses co-infecting a hypovirulent isolate of the plant pathogenic fungus Sclerotinia sclerotiorum. Virology 428, 77–85. https://doi.org/10.1016/j.virol.2012.03.015
- Xu Z, Wu S, Liu L, Cheng J, Fu Y, Jiang D, Xie J (2015): A mitovirus related to plant mitochondrial gene confers hypovirulence on the phytopathogenic fungus Sclerotinia sclerotiorum. Virus Res. 197, 127–136. https://doi.org/10.1016/j.virusres.2014.12.023