Virus Research 263 (2019) 169–172

Contents lists available at ScienceDirect

Virus Research

journal homepage: www.elsevier.com/locate/virusres

Short communication

Molecular characterisation of a novel gemycircularvirus associated with olive trees in Italy

Michela Chiumenti[a], Claudia Greco[a,b], Ilaria Antelmi[b], Valentina Sion[b], Giuseppe Altamura[a], Franco Nigro[b], Pasquale Saldarelli[a,*]

[a] Istituto per la Protezione Sostenibile delle Piante, Consiglio Nazionale delle Ricerche, Via Amendola, 122/D 70126 Bari, Italy
[b] Dipartimento di Scienze del Suolo, della Pianta e degli Alimenti, Università degli studi di Bari "Aldo Moro", Via Amendola, 165/A 70126 Bari Italy

ARTICLE INFO

Keywords:
Olive
Rolling circle amplification
Circular ssDNA
Gemycircularvirus

ABSTRACT

A new gemycircularvirus sequence was obtained from total DNAs extracted from olive leaves and twigs tissues. Its complete genome consists of a single-stranded circular DNA of 2,145 nt, whose sequence was obtained by rolling circle amplification (RCA). Genome analysis identified three open reading frames, sharing homologies with the coat and replication-associated proteins, these latter in the anti-parallel strand, of known gemycircularvirus species. Search for homologies showed that the circular ssDNA sequence is distantly related to other gemycircularviruses thus originating from a new virus species, for which the name olive-associated gemycircularvirus 1 (OaGV1) is proposed. A survey in two different olive-growing areas of the Apulian region (Southern Italy), showed a limited distribution of OaGV1.

In 2016, the International Committee on Taxonomy of Viruses (ICTV) created a new family of single-stranded circular DNA viruses, the *Genomoviridae*, which contains a single genus, *Gemycircularvirus* (Krupovic et al., 2016). Members of this family present icosahedral and non-enveloped capsids 20–22 nm in diameter (Yu et al., 2010; Sikorski et al., 2013). Their circular single-stranded DNA genomes, ranging from 2.1 to 2.3 kb, mainly consist of two proteins, a coat protein (CP) and a spliced replication initiation protein (Rep). A conserved stem-loop structure, including the putative origin of DNA replication is located between the 5′ ends of the two main ORFs, which are therefore anti-parallel in the two opposite reading frames.

The first report of a virus belonging to this recently defined genus dates back to 2010, when Yu et al. described a mycovirus able to induce hypovirulence on infected *Sclerotinia sclerotiorum* population (Sclerotinia sclerotiorum hypovirulence-associated DNA virus 1, SsHADV1) and to establish infection when applied extracellularly as purified virus particles. Since then, these viruses have been reported to be associated with hosts other than fungi, and belonging to different kingdoms (Yu et al., 2010), such as animals (Rosario et al., 2012; Dayaram et al., 2015; Li et al., 2015), including human beings (Lamberto et al., 2014; Phan et al., 2015; Halary et al., 2016) and plants (Dayaram et al., 2012; Kraberger et al., 2015; Male et al., 2015; Marzano and Domier, 2016). Moreover, they were also found in environmental samples like sewage, faeces, and water (Sikorski et al., 2013; Conceicao-Neto et al., 2015; da Silva Assis et al., 2016).

In the last years, the application to viral metagenomics of sequence unbiased techniques, i.e. high-throughput sequencing, and non-specific rolling circle amplification (RCA), has increased the number and facilitated the discovery of novel viruses in various ecosystems in particular those related to the newly described family of *Genomoviridae* (Delwart, 2007; Pantaleo and Chiumenti, 2018). The application of RCA allowed the identification and sequencing of the first gemycircularvirus associated to olive (*Olea europaea* L.) in Apulia (Italy). In the present study, we report the complete genome description and molecular characterization of a new gemycircularvirus provisionally named olive-associated gemycircularvirus 1 (OaGV1). In addition, an extended survey was conducted in southern Apulia olive orchards to investigate its distribution in orchards and intraspecific variability.

Olive samples were collected in spring 2015 from plants located at Palagiano, in the province of Taranto (Apulia region, Southern Italy). Trees in this experimental orchard were obtained by somatic embryogenesis. Total DNA was extracted using a CTAB protocol from 200 mg of petioles and leaves of single trees as previously described (Murray and Thompson, 1980). Total DNA extracts from three plants were used in a random priming rolling circle amplification RCA, using TempliPhi (GE Healthcare). One µl of rolling circle amplified product was digested with BamHI, SpeI, SalI and HindIII. One (L1) out of three samples analyzed, yielded a linear DNA fragment of about 2.2 kb when digested

* Corresponding author.
E-mail address: pasquale.saldarelli@ipsp.cnr.it (P. Saldarelli).

https://doi.org/10.1016/j.virusres.2019.01.016
Received 12 October 2018; Received in revised form 21 January 2019; Accepted 26 January 2019
Available online 29 January 2019
0168-1702/

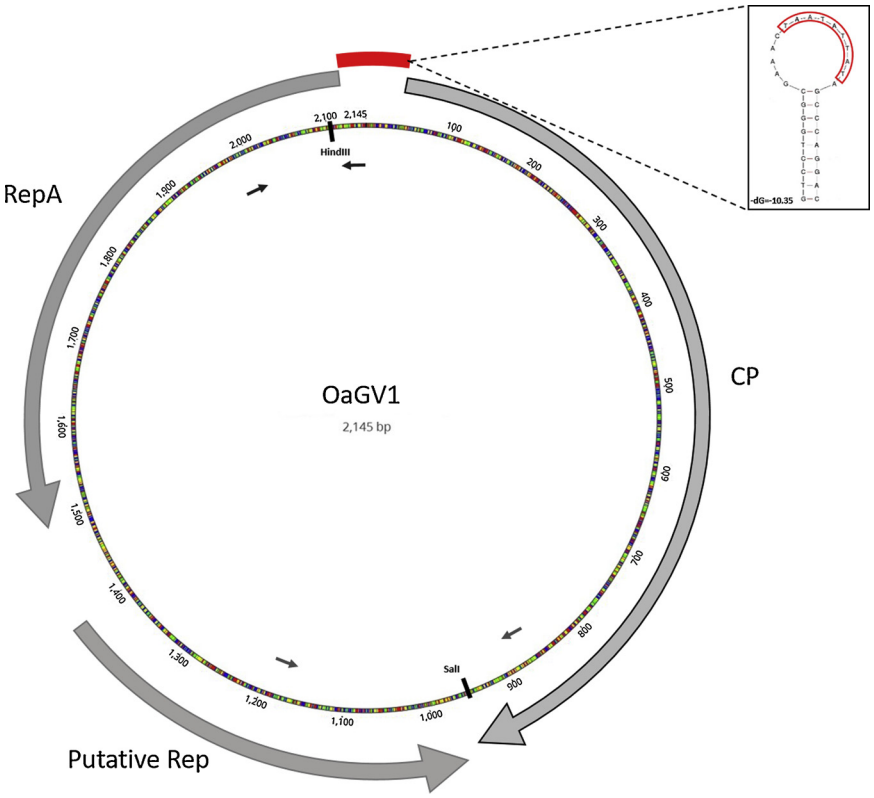

Fig. 1. Graphical representation of the OaGV1 genome organization. Grey bold arrows indicate the three open reading frames. The red bar stands for the long intergenic region, which can fold producing a hairpin structure (in the upper right box with its free energy, while red lines frame the putative origin of replication). Small black arrows localize the primers used for the survey. The small black vertical bars on the genome point to the restriction site positions.

with SalI and HindIII. The restriction-digested DNA fragments were ligated to SalI and HindIII-digested pUC19 (Fermentas,USA), respectively (Supplementary Fig. 1).

Two independent recombinant plasmids harbouring SalI or HindIII RCA-digested fragments were sequenced from both ends. All sequences, once submitted to Blast search, matched genomic DNAs from genomo-like entities. Cap contig assembly using Bioedit showed that all the fragments were perfectly overlapping to long extends and produced a single-stranded circular DNA of 2145 nt in length with a G + C content of 52.17%. Sequence analysis of this circular DNA for potential coding proteins, identified three main open reading frames (ORF) (Fig. 1). Blast analysis of the full genome found that the closest viruses are odonata associated gemycircularvirus 1 isolate OdaGmV-1-US-260BC-12 (OdaGmV1; Acc. Nr. KM598385) and giant panda associated gemycircularvirus strain gpge006 (GpGmCV; Acc. Nr. MF327563) with 65.27 and 41.14% identical nucleotides, respectively.

The largest ORF is in positions 97–975 and reading frame +1. It starts with an ATG codon, yielding a protein of 292 amino acids (aa) having a 32.85 kDa molecular weight. Blast analysis of this sequence found no significant similarity at the nucleotide level, while, using the Blastx algorithm, a significant degree of similarity was retrieved with the putative coat proteins (CP) of blackbird associated gemycircularvirus 1 (BlBaGV1; Accession nr. YP_009109740; 81% coverage, maximum aa identity of 59%, E value 1e-88), dragonfly-associated circular virus 2 (DfasCV2; Accession nr. YP_009021855; 81% coverage, maximum aa identity of 57%, E value 7e-83), and SsHADV1 (Accession nr. AGP05336; 80% coverage, maximum aa identity of 57%, E value 1e-82). On the complementary strand, a second ORF starts with an ATG at position 2,120 to 1521 with reading frame -2. It produces a putative protein of 199 aa in length and 22.47 kDa in weight. Blastx search for homologies finds similarity with the replication protein (RepA) of OdaGmV1 (Accession nr. AJD07461; 99% coverage, maximum aa identity of 80%, E value 7e-105) and DfasCV2 (Accession nr. YP_009021858; 99% coverage, maximum aa identity of 67%, E value 3e-86). A third ORF starts at position 396 until nt 1073. It is read in frame -3, and produces a protein of 107 aa and 12.25 kDa, with the Blastx highly significance to replication-associated protein of GpGmCV (Accession nr. ASH99154; 88% coverage, maximum aa identity of 94%, E value 7e-59) and DfasCV2 (Accession nr. YP_009021856; 89% coverage, maximum aa identity of 90%, E value 4e-57).

Between the RepA and the CP ORFs there is a 122-nt long intergenic region (LIR), which, as in other gemycircularviruses, is nine nucleotide (TAATATTAT) long. This sequence is considered the origin of viral genome replication, by analogy with the geminiviruses (ori; Fig. 1). A putative TATA box (TATATA), 30 nucleotides downstream the non-anucleotide and immediately upstream of a CG-rich region, known to be involved in transcriptional regulation (Eagle and Hanley-Bowdoin, 1997; Fenoll et al., 1988, 1990), was also identified. In addition, a short intergenic region (SIR) extends for 98 nt, from the end of the CP ORF to the end of the replication-associated protein.

Sequence comparison at nucleotide level was performed through Sequence Demarcation Tool (SDT v1.2; Muhire et al., 2014) using the ClustalW alignment algorithm (Thompson et al., 2003). All the gemycircularvirus genomes deposited in GenBank RefSeq database together with OaGV1 (https://www.ncbi.nlm.nih.gov/refseq/) were pairwise aligned to evaluate the percentage of sequence identity. As showed in Supplementary Table 1, OaGV1 revealed a percentage of sequence identity below 65% with all the other described gemycircularviruses described, except for dragonfly-associated circular virus 2 (NC_023871), with which it shares 68.4% of nucleotide identity.

A phylogenetic analysis on the CP proteins was conducted in MEGA7 (Kumar et al., 2016) using the evolutionary model proposed by Krupovic et al. (2016) including the closely related species retrieved by Blast search and few other species associated with plants. Surprisingly, OaGV1 appears to belong to the clade of the BlBaGV1, SsHADV1, OdaGmV1 and DfasCV2, while the other plant-associated gemycircularviruses form a different group (Fig. 2).

Two specific set of primers selected in the surroundings of the restriction sites used for RCA digestion (see Table 1, HindIII-for/HindIII-rev and SalI-for/SalI-rev) were designed in order to evaluate the

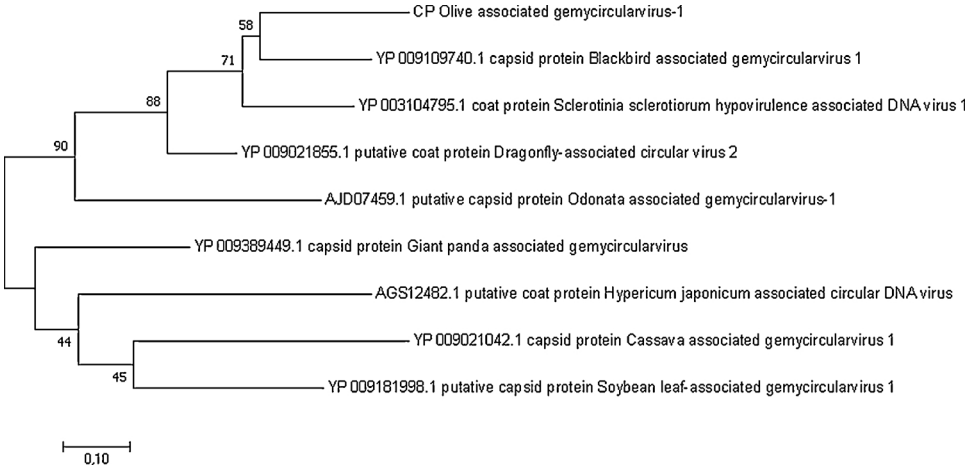

Fig. 2. Unrooted molecular phylogenetic tree obtained with Maximum Likelihood method using the evolutionary model described by Krupovic et al. (2016). Protein sequences used for phylogenetic analysis, were aligned using Bioedit (Hall, 1999) and manually edited to remove gaps. All viruses are indicated with their GenBank accession numbers followed by their full names.

Table 1
Primers designed on OaGV1 genome around the considered restriction sites and used for the survey and for variability study. Names, 5′- > 3′ sequence, the location on the genome and amplicon size are indicated for each primers.

Primer name	Sequence 5' - > 3'	Location on the genome	Amplicon size
HndIII_for	GCTGCTCTCTTGCAATGATGCAC	1991–2013	231
HndIII _rev	GGTACGCCATCGGACGGCA	106–88	
SalI _for	CCTGGTGTATGATGACGACGAGTC	861–884	358
SalI _rev	GGAGTGGTTAGGGGCTCAGGC	1217–1197	

Table 2
Intraspecific variability: The emimatrices below show the percentage of sequence identity among amplicons obtained in the two different region under study. Matrix A represents the amplicons obtained with HindIII for/rev primers and matrix B amplicons obtained using SalI for/rev primers. Alignments were run using ClustalW algorithm using Sequence Demarcation Tool (SDTv1.2).

A. identity matrix HindIII for/ HindIII rev primers

HindIII				
L1	100.00			
14	100.00	100.00		
L3	97.31	97.31	100.00	
12B	100.00	100.00	97.31	100.00

B. identity matrix SalI for/ SalI rev primers

SalI				
L1	100.00			
14	95.68	100.00		
L3	97.83	93.83	100.00	
12B	99.69	96.30	97.53	100.00

presence of this virus in the olive orchards and investigate its sequence variability. In spring 2017, ninety olive plants were sampled in southern Apulia, 41 and 49 from Lecce and Taranto provinces, respectively. Among these trees, 18 different cultivars and one seedling were sampled (Supplementary Table 2). Total DNA extracts obtained from leaves and petioles using the above-mentioned CTAB-based protocol, were analyzed with SalI-for/SalI-rev set of primers by PCR. Only three samples from the same area (Taranto province) of the first positive tree were positive for OaGV1. Sequences from OaGV1-infected trees were further analysed in the HindIII-for/HindIII-rev amplified regions, in order to gather more data about intraspecific variability. Amplicons of the expected size were obtained from all plants and were sequenced from both ends.

The alignment of the obtained amplicons, including the fragment from the original L1 tree, disclosed a percentage of nucleotide identity ranging between 97.31% to 100% in the region amplified with HindIII-for/HindIII-rev primers, and 93.83% to 99.69% in that amplified with the SalI-for/SalI-rev primers (Table 2). The higher values of sequence identity obtained with HindIII-for/HindIII-rev confirm the importance and conservation of this region, since it contains a fragment of the LIR region including the putative origin of replication and the hairpin structure (see Fig. 1). On the other hand, amplicons in the SalI-for/SalI-rev region show a slightly higher variability as it includes the SIR and part of the CP and putative Rep protein, which are likely more tolerant to nucleotide variation.

To date viruses belonging to the family *Genomoviridae* have never been reported as plant pathogens. Indeed, when found in plants they were considered as viruses likely infecting fungi associated with plant species under analysis and, in our study, such hypothesis was possibly supported by the presence of SsHADV1 in the same phylogenetic group of OaGV1. This prompted to screen the total DNAs from OaGV1-positive olive trees with broad-spectrum panfungal primers. The Internal Transcribed Spacer (ITS) region, partial sequence of β-tubulin (TUB2), and Actin (ACT) genes were amplified using the primers ITS5/4 (White et al., 1990), T1/bt2b (O'Donnell and Cigelnik, 1997; Glass and Donaldson, 1995), and ACT-512 F/ACT-986R (Carbone and Kohn, 1999), respectively. The expected amplicons about 700, 600 and 300 bp for ITS, TUB2, and ACT, respectively, were cloned and sequenced from both ends. Sequence analysis of the amplicons showed similarity with a diversity of fungi belonging to a range of genera, including *Pleospora*, *Hormonema*, *Arthrocaldium*, *Dactylonectria*, *Isaria*, *Toxicocladosporium*, *Alternaria*, and *Mycosphaerella*. The wide range of fungal genera retrieved does not suggest an unambiguous fungal group as host of OaGV1. Although it is probable that OaGV1 is a mycovirus, further efforts for host range identification and characterization are needed, like synthesizing and using infectious clones of this virus for pathogenesis tests on selected fungal species.

Olive and its related agricultural context represent an additional environment in which a new gemicyrcularvirus species was found. As in other members of the genus an almost total absence of information related to the host range, biology and epidemiology (except for SsHADV-1) exist, a gap that needs to be closed. OaGV1 showed the closest nucleotide sequence identity (68.4%) with dragonfly-associated circularvirus 2, which is below the thresholds set by ICTV in the family *Geminiviridae*, ranging from 75% for mastreviruses to 89% for curtoviruses and begomoviruses, while no species demarcation criteria exist yet for the genus *Gemycircularvirus*. Altogether, these findings endorse the proposal that this is a new species, tentatively named olive-associated gemycircularvirus 1 (OaGV1). The sequence was deposited in Genbank under accession number MH444690.

Genbank accession

MH444690.

Conflict of interest

The authors MC, CG, IA, VS, GA, FN and PS declare that they have no conflict of interest.

Ethical approval

This article does not contain any studies with human participants or animals performed by any of the authors.

Appendix A. Supplementary data

Supplementary material related to this article can be found, in the online version, at doi:https://doi.org/10.1016/j.virusres.2019.01.016.

References

Carbone, I., Kohn, L.M., 1999. A method for designing primer sets for speciation studies in filamentous ascomycetes. Mycologia 91, 553–556.

Conceicao-Neto, N., Zeller, M., Heylen, E., Lefrere, H., Mesquita, J.R., Matthijnssens, J., 2015. Fecal virome analysis of three carnivores reveals a novel nodavirus and multiple gemycircularviruses. Virol. J. 12–79.

da Silva Assis, M.R., Vieira, C.B., Fioretti, J.M., Rocha, M.S., de Almeida, P.I.N., Miagostovich, M.P., Fumian, T.M., 2016. Detection and molecular characterization of gemycircularvirus from environmental samples in Brazil. Food Environ. Virol. 8 (4), 305–309.

Dayaram, A., Opong, A., Jaschke, A., Hadfield, J., Baschiera, M., Dobson, R.C., Offei, S.K., Shepherd, D.N., Martin, D.P., Varsani, A., 2012. Molecular characterisation of a novel cassava associated circular ssDNA virus. Virus Res. 166, 130–135.

Dayaram, A., Potter, K.A., Pailes, R., Marinov, M., Rosenstein, D.D., Varsani, A., 2015. Identification of diverse circular single-stranded DNA viruses in adult dragonflies and damselflies (*Insecta*: *Odonata*) of Arizona and Oklahoma, USA. Infect. Genet. Evol. 30, 278–287.

Delwart, E.L., 2007. Viral metagenomics. Rev. Med. Virol. 17 (2), 115–131.

Eagle, P.A., Hanley-Bowdoin, L., 1997. Cis elements that contribute to geminivirus transcriptional regulation and the efficiency of DNA replication. J. Virol. 71 (9), 6947–6955.

Fenoll, C., Black, D.M., Howell, S.H., 1988. The intergenic region of maize streak virus contains promoter elements involved in rightward transcription of the viral genome. EMBO J. 7 (6), 1589–1596.

Fenoll, C., Schwarz, J.J., Black, D.M., Schneider, M., Howell, S.H., 1990. The intergenic region of maize streak virus contains a GC-rich element that activates rightward transcription and binds maize nuclear factors. Plant. Mol. Biol. 15 (6), 865–877.

Glass, N.L., Donaldson, G.C., 1995. Development of primer sets designed for use with the PCR to amplify conserved genes from filamentous infection due to *Phaeoacremonium* spp. J. Clin. Microbiol. 41, 1332–1336.

Halary, S., Duraisamy, R., Fancello, L., Monteil-Bouchard, S., Jardot, P., Biagini, P., Gouriet, F., Raoult, D., Desnues, C., 2016. Novel single-stranded DNA circular viruses in pericardial fluid of patient with recurrent pericarditis. Emerg. Infect. Dis. 22 (10), 1839.

Hall, T.A., 1999. BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. Nucl. Acids. Symp. Ser. 41, 95–98.

Kraberger, S., Farkas, K., Bernardo, P., Booker, C., Arguello-Astorga, G.R., Mesleard, F., Martin, D.P., Roumagnac, P., Varsani, A., 2015. Identification of novel Bromus- and Trifolium-associated circular DNA viruses. Arch. Virol. 160, 1303–1311.

Krupovic, M., Ghabrial, S.A., Jiang, D., Varsani, A., 2016. Genomoviridae: a new family of widespread single-stranded DNA viruses. Arch. Virol. 161 (9), 2633–2643.

Kumar, S., Stecher, G., Tamura, K., 2016. MEGA7: molecular evolutionary genetics analysis version 7.0 for bigger datasets. Mol. Biol. Evol. 33, 1870–1874.

Lamberto, I., Gunst, K., Muller, H., Zur Hausen, H., de Villiers, E.M., 2014. Mycovirus-like DNA virus sequences from cattle serum and human brain and serum samples from multiple sclerosis patients. Genome Announc. 2 e00848–00814.

Li, W., Gu, Y., Shen, Q., Yang, S., Wang, X., Wan, Y., Zhang, W., 2015. A novel gemycircularvirus from experimental rats. Virus Genes 51, 302–305.

Male, M.F., Kami, V., Kraberger, S., Varsani, A., 2015. Genome sequences of *Poaceae*-associated Gemycircularviruses from the Pacific Ocean Island of Tonga. Genome Announc. 3 e01144–01115.

Marzano, S.Y.L., Domier, L.L., 2016. Novel mycoviruses discovered from metatranscriptomics survey of soybean phyllosphere phytobiomes. Virus. Res. 213, 332–342.

Muhire, B.M., Varsani, A., Martin, D.P., 2014. SDT: a virus classification tool based on pairwise sequence alignment and identity calculation. PLoS ONE 9 (9), e108277. https://doi.org/10.1371/journal.pone.0108277.

Murray, M.G., Thompson, W.F., 1980. Rapid isolation of high molecular weight plant DNA. Nucleic. Acids. Res. 8 (19), 4321–4326.

O'Donnell, K., Cigelnik, E., 1997. Two divergent intragenomic rDNA ITS2 types within a monophyletic lineage of the fungus *Fusarium* are nonorthologous. Mol. Phylogenet. Evol. 7, 103–116.

Pantaleo, V., Chiumenti, M., 2018. Viral Metagenomics – Methods and Protocols, first ed. Humana Press, New York.

Phan, T.G., Mori, D., Deng, X., Rajindrajith, S., Ranawaka, U., Fan Ng, T.F., Bucardo-Rivera, F., Orlandi, P., Ahmed, K., Delwart, E., 2015. Small circular single stranded DNA viral genomes in unexplained cases of human encephalitis, diarrhea, and in untreated sewage. Virology 482, 98–104.

Rosario, K., Dayaram, A., Marinov, M., Ware, J., Kraberger, S., Stainton, D., Breitbart, M., Varsani, A., 2012. Diverse circular ssDNA viruses discovered in dragonflies (*Odonata*: *Epiprocta*). J. Gen. Virol. 93, 2668–2681.

Sikorski, A., Massaro, M., Kraberger, S., Young, L.M., Smalley, D., Martin, D.P., Varsani, A., 2013. Novel myco-like DNA viruses discovered in the faecal matter of various animals. Virus Res. 177, 209–216.

Thompson, J.D., Gibson, T.J., Higgins, D.G., 2003. Multiple sequence alignment using ClustalW and ClustalX. Curr. Protocols Bioinform. 1, 2–3.

White, T.J., Burns, T., Lee, S., Taylor, J., 1990. Amplification and direct sequencing of fungal ribosomal RNA genes for phylogenetics. In: Innis, M.A., Gelfand, D.H., Sninsky, J.J., White, T.J. (Eds.), PCR Protocols: a Guide to Methods and Applications. Academic Press, San Diego, pp. pp. 315–322.

Yu, X., Li, B., Fu, Y., Jiang, D., Ghabrial, S.A., Li, G., Peng, Y., Xie, J., Cheng, J., Huang, J., Yi, X., 2010. A geminivirus-related DNA mycovirus that confers hypovirulence to a plant pathogenic fungus. Proc. Natl. Acad. Sci. U. S. A. 107, 8387–8392.