First detection of mycoviruses in Gnomoniopsis castaneae suggests a putative horizontal gene transfer event between negative-sense and double-strand RNA viruses

UNIVERSITÀ DEGLI STUDI DI TORINO

First detection of mycoviruses in *Gnomoniopsis castaneae* suggests a putative horizontal gene transfer event between negative-sense and double-strand RNA viruses

Franco Ferilli[1,3], Guglielmo Lione[1], Paolo Gonthier[1], Massimo Turina[2], Marco Forgia[2*].

1: University of Torino, Department of Agricultural, Forest and Food Sciences (DISAFA), Largo Paolo Braccini 2, 10095 Grugliasco, Torino, Italy.

2: Institute for Sustainable Plant Protection, National Research Council of Italy, Strada delle Cacce, 73, Torino 10135, Italy.

3: currently an EFSA staff member in the Environment, Plants & Ecotoxicology Unit, European Food Safety Authority (EFSA), via Carlo Magno 1A, 43126 Parma, Italy

*Corresponding author. E-mail: marco.forgia@ipsp.cnr.it

Keywords

Gnomoniopsis castaneae, mycovirus, HGT, mitovirus, chrysovirus, Negarnaviricota.

Abstract

Gnomoniopsis castaneae is an ascomycetous fungus mainly known as a major pathogen of chestnut causing nut rots, although it is often found as an endophyte in chestnut tissues. To date, no virus has been reported as associated with to this fungus. Here, a collection of *G. castaneae* isolates from several European countries was screened to detect mycoviruses infecting the fungus: for the first time we report the identification and prevalence of mitovirus Gnomoniopsis castaneae mitovirus 1 (GcMV1) and the chrysovirus Gnomoniopsis castaneae chrysovirus 1 (GcCV1). Interestingly, we provide evidence supporting a putative horizontal gene transfer between members of the phyla Negarnaviricota and Duplornaviricota: a small putative protein of unknown function encoded on the RNA3 of GcCV1 (*Chrysoviridae*) has homologues in the genome of viruses of the family *Mymonaviridae*.

Highlights

- The first two viruses infecting *Gnomoniopsis castaneae* were detected and characterized making them a suitable model to study possible effects on the virulence of the host.
- A collection of fungal isolates from different European regions was screened and the identified viruses were present in a broad area around Italy, Swiss and Czech Republic.
- The characterization of Gnomoniopsis castaneae chrysovirus 1 showed a small ORF encoding for a protein of unknown function with homologs in some chrysoviruses and in members of the order *Mononegavirales,* suggesting an HGT event between distantly related RNA viruses.

Introduction

Since the initial confirmation of a fungal infection by a mycovirus in the edible fungus *Agaricus bisporus* (Ghabrial et al., 2015; Hollings 1962), extensive research has been dedicated to investigating fungal viruses and the impact on their hosts. To date, viral infections have been observed across all major fungal groups; with most of the identified mycoviruses possessing RNA genomes, comprising positive (+) or negative (-) sense single-stranded RNA (ssRNA), or double-stranded RNA (dsRNA). While there are mycoviruses with ssDNA genomes, they are

significantly less prevalent (Kondo, et al., 2022). In recent years, the possibility of employing viruses as biocontrol agents against fungal plant pathogens has gained attention (Wagemans et al., 2022).

Mycoviruses do not have an extracellular phase, meaning their transmission is confined to intracellular content exchange mechanisms (Voth et al., 2006). The two primary transmission pathways are horizontal transfer, which occurs through cytoplasmic attachment of hyphae within the same fungal species (anastomosis), and vertical transmission through asexual spore production (Pearsonet al., 2009; Xie et al., 2014). Mycoviral infections are typically persistent within their hosts without significantly altering their phenotypes. Nevertheless, a multitude of mycoviruses have been found to induce either hypo- or hypervirulent effects on their fungal hosts. In cases of hypovirulence, the infection results in a reduction of the host's virulence, whereas hypervirulent viral infections amplify the pathogenic effects of the host fungus (Kotta-Loizou, 2021).The best studied example of mycoviruses that confer hypovirulence on plant pathogenic fungi is the hypovirus of the chestnut blight fungus *Cryphonectria parasitica* (Choi et al., 1992; Nuss, 2005). Nevertheless, recent studies changed our perspective on the use of mycoviruses as biocontrol agents showing that virus-induced hypovirulence could be linked to a change in the fungus lifestyle from pathogenic to endophytic, thereby protecting the plant through eliciting the plant immune response to pathogens (Tian et al., 2020; Zhou et al., 2021). Mycoviruses also have the capacity to induce a wide range of phenotypes in their fungal hosts (other than hypovirulence), some of which are difficult to evaluate due to the hidden role played by viruses on hosts grown on artificial substrates in axenic cultures: their biological significance is more often related to complex interaction between the infected fungus and the environment (Chun et al., 2018; Liu et al., 2019; Espino-Vázquez et al., 2020). Furthermore, the high frequency of multiple infections observed in fungi is a major problem when evaluating the possible phenotype induced by one mycovirus (Picarelli et al., 2019; Deakin et al., 2017).

In this study, we investigated viral infections of the fungus *Gnomoniopsis castaneae* (syn. *G. smithogilvyi*), a fungus with alternative lifestyles oscillating between endophytic and phytopathogenic (Lione et al., 2019). Since the early 2000s, chestnut growers in various countries, including Italy, Switzerland, France, New Zealand, and Australia, have reported a significant increase in chestnut fruit rot incidence, displaying peculiar symptoms not clearly associated with known fungal pathogens or pests affecting chestnuts. In 2012, the agent responsible for the chestnut fruit rot was identified as a new pathogenic fungus and named *G. castaneae* (Lione et al., 2019; Visentin et al., 2012).Infected chestnuts fruits exhibit symptoms including a chalk-white interior with severe dehydration or rotting appearance with a dark brown color (Lema et al., 2023). On the chestnut burrs, the pathogen produces minute fruiting bodies (perithecia) containing sexual spores, which are the result of the fungal reproduction. Asexual fruiting bodies (acervuli) leading to the production of asexual spores (conidia) can also be observed on burrs or rotten fruits. The fungus infects by means of spores, both sexual and asexual, penetrating through the flowers of chestnut trees (Sillo et al., 2017). Currently, *G. castaneae* is considered a significant global threat to chestnut cultivation, having expanded its distribution across both hemispheres, encompassing many of the areas where chestnuts are grown (see https://gd.eppo.int/taxon/GNMPCA for current distribution of *G. castaneae*) (Lione et al., 2016). *G. castaneae* is a very adaptable fungus and presents a formidable challenge in terms of monitoring and impact assessment. Not only it is the primary cause of fruit rot, but it has also been observed in association with bark cankers and leaf necrosis, affecting not only chestnuts but also other species such as boxwood and hazel. Infected chestnuts maintain an outwardly healthy appearance, while cankers on chestnut trees bear a striking resemblance to those caused by *C. parasitica*. Likewise, leaf necrosis does not provide definitive clues for diagnosis. Moreover, *G. castaneae* acts as an endophyte within the chestnut tree, residing in the plant's green tissues without triggering visible symptoms. This unique characteristic renders its detection and diagnosis a challenge (Lione et al., 2019).

Since the infection biology, ecology and epidemiology of *G. castaneae* are not fully understood, all contributing factors and variables, such as different mycoviruses hosted by different fungal isolates, have to be taken into consideration. Understanding a possible hypo- or hypervirulence caused by the occurring mycoviruses might

help to investigate potential applications in biological control and disease management. Thus, the research on mycoviruses could be an important component in the development of strategies to mitigate complex plant diseases such as those caused by *G. castaneae*, where a direct correlation between presence and symptom development is not always present. Therefore, the main objective of this study is to give a first insight into the fungal virome of the emerging and destructive plant pathogen *G. castaneae*. Here, we report on the identification of two new viruses, tentatively named Gnomoniopsis castaneae mitovirus 1 (GcMV1) and Gnomoniopsis castaneae chrysovirus 1 (GcCV1). This is the first report of viruses infecting a member of the fungal genus *Gnomoniopsis.* In addition, further analysis of the proteins encoded by GsCV1 revealed hints on some domain conservation among proteins present in some chrysovirids (Phylum Duplornaviricota) and uncharacterized proteins occasionally present in mymonavirids (Phylum Negarnaviricota), possibly caused by horizontal gene transferring events. This study provides the basis for further investigations of the impact of GcMV1 and GcCV1 on *G. castaneae*.

Material and methods

Fungal isolates and their maintenance

G. castaneae isolates used in this work are part of the collection of the Forest Pathology Laboratory of the University of Turin. The origin of isolates is available in Supp. Table 1. Isolates were obtained in different years from nuts, cankers and green plant tissues as previously described (Lione et al. 2016) and were identified based on the macro- and micromorphological features of pure cultures and using taxon specific primers (Lione et al., 2015). Isolates were maintained on PDA medium and cultivated in liquid PDB media to obtain mycelia suitable for RNA extraction.

High throughput sequencing and bioinformatic analysis

A subset consisting of 30 isolates of the complete fungal collection was selected for high throughput sequencing (HTS) of the total RNAs to detect evidence of viral infection. This subset of isolates was selected taking into consideration the part of the plant that was sampled and geographic origin. The list of isolates included in this subset is available in Supp. Table 1 (Isolates number 1 to 30 based on the population genetics study from Sillo et al., 2017). Lyophilized fungal mycelia was homogenized using a bead beater (MP Biomedicals, Fisher Scientific, Hampton, NH, USA) and 0.2 mm diameter glass beads in a 2 ml O-ring tube. Total RNA from the fungal mycelia was extracted using a Spectrum Plant Total RNA Kit (Sigma-Aldrich, St. Louis, MO, US) following the manufacturer's instructions. RNA concentration and quality was checked using a NanoDrop 2000 Spectrophotometer (Thermo Fisher Scientific, Waltham, MA, USA) and samples were pooled in a single tube to perform HTS analysis. Ribosomal RNA depletion and library construction were made using TruSeq Stranded Total RNA Ribo-Zero H/M/R Gold kit (Illumina, San Diego, CA, USA); Illumina sequencing was performed with a NovaSeq 6000 platform (Illumina, San Diego, CA, USA). Ribosomal depletion, library preparation and illumina sequencing were provided by Macrogen (Seoul, Republic of Korea). The obtained reads were cleaned following a pipeline relying on BBTools (dx.doi.org/10.17504/protocols.io.gydbxs6) and assembled using Trinity version 2.9.1. Viral contigs were identified in the metatranscriptome by performing a BLASTP search of the obtained transcripts against the National Center for Biotechnology Information (NCBI) non redundant (nr) database with DIAMOND as described previously (Forgia et al., 2022). The viral contigs were retrieved from the BLASTP result with a manual selection. ORF prediction and conserved protein domains identification in the putative proteins were performed using the NCBI tool ORF Finder (https ://www.ncbi.nlm.nih.gov/orffinder/) and conserved domain search (https://www.ncbi.nlm.nih.gov/Structure/cdd/wrpsb.cgi). Mappings of reads on the viral contigs were made using Bowtie2 (Langmead et al., 2012) software, and results were retrieved from the generated SAM file with SamTools (Li et al., 2009) and displayed with Tablet (Milne et al., 2013).

Protein identity and structure conservation was investigated through alignment of the proteins of interest using MAFFT (Katoh et al., 2019); the obtained alignment was used to generate an identity matrix using Discovery studio software, and graph were generated with R (R Developement Core Team, 2010) using the libraries reshape2 and ggplot2 (Wickham, 2007). Structure modeling of the protein of interest was made with Alphafold2 (Jumper et al., 2021) using the ColabFold (Mirdita et al., 2022). Comparisons among structures were performed with ChimeraX (Pettersen et al. 2021) obtaining the root-mean-spuare distance (RMSD) for the portion of the structure that is conserved (Carugo et al. 2001).

Molecular analysis

To connect each viral sequence identified to the specific infected fungal isolate present in the pooled Illumina sample, quantitative RT-PCR (qRT-PCR) was performed using primers designed based on the viral contigs. qRT-PCR was performed not only on each individual sample included in the pooled RNAs used for HTS but also on the remaining isolates used in this work showing the mean cycle threshold (Ct) value between two PCR repetition. (Supp. Table 1). qRT-PCRs were made in 10 µl total volume using a 2X Bio-Rad SYBR Green Master Mix in a BioRad CFX Connect Real-Time PCR Detection System (BioRad, Hercules, CA, USA). A list of the primers used in this work is available in Supp. Table 2.

Rapid amplification of cDNA ends (RACE) analysis was performed on the 5' and 3’ end of each viral sequence to confirm the *in silico* assembly obtained. The RACE method used (Rastgou et al., 2009) rely on the synthesis of cDNA using specific primers located close to the 5’ and 3’ ends of the contigs assembled in silico (Supp. Table 2). cDNAs were synthesized using Superscript IV (Thermo Fisher Scientific, Waltham, MA, USA) reverse transcriptase and tagged with polyA or polyG using deoxynucleotidyl transferase (Promega, Madison, WI, USA). PCRs were then performed using primer complementary to the polyA or polyG added, and a specific primer on the viral sequence (Supp. Table 2). Obtained PCR bands were purified from electrophoresis gel using Zymo gel DNA recovery kit (Zymo Research, Irvine, CA, USA), inserted in a plasmid using the pGEMT easy vector kit (Promega, Madison, WI, USA) and cloned in *E. coli* DH5α cells. Transformed colonies were selected and obtained plasmids were purified using ZR Plasmid Miniprep kit (Zymo Research, Irvine, CA, USA) and sent for Sanger sequencing to Biofab srl (Rome, Italy). Each nucleotide in the sequence was confirmed by at least three clones. For GcCV1 RNA4, no clones could be obtained for the 5’ end. In this case we used a protocol relying on the ligation of a blocked adaptor at the 3’ end of the RNA sample that was described in detail in a previous work (Forgia et al. 2022). Briefly, after the ligation of the blocked adaptor to the 3’ of the RNA sample, cDNA was produced using a primer complementary to the adaptor and PCR bands were obtained using a specific primer with the primer complementary to the adaptor (Supp. Table 2).

Phylogenetic analysis

Phylogenetic analysis was performed using the putative RdRp from GcMV1 and GcCV1 for building Maximum likelihood (ML) trees. Alignment of the putative RdRps was made using MAFFT through the web server interface and the obtained alignments were submitted to IqTree webserver for building the ML Phylogenetic trees using the automatic model selection option and the ultrafast bootstrap method (with 1000 replicates) for calculating the statistical robustness of the clades obtained. The sequences included in the phylogenetic tree of the family *Chrysoviridae* were retrieved from the ICTV resources (https://ictv.global/report/chapter/chrysoviridae/chrysoviridae/resources) using members of the genus *Totivirus* as outgroup. The sequences included in the phylogenetic tree of the family *Mitoviridae* were obtained from the latest ICTV proposal (https://ictv.global/taxonomy/taxondetails?taxnode_id=202213793), including members of the family *Narnavidae* as outgroup, the RdRp sequence of Sclerotinia sclerotiorum mitovirus 15 was also included in the analysis as the closest sequence to the RdRp of GcMV1 (Supp. Table 3). For the Mononegavirales related phylogenetic tree, the sequences included were retrieved from Pagnoni study (Pagnoni

et al. 2023) adding all the Mononegavirales-related sequences encoding for ORFD-like putative proteins listed in Supp. Table 4.

Results

Molecular characterization of Gnomoniopsis castaneae mitovirus 1

From the NGS analysis of the pooled sample of *G. castaneae,* one contig related to viruses belonging to the family *Mitoviridae* was identified. Its complete sequence was reconstructed with RACE analysis and the obtained sequence was named Gnomoniopsis castaneae mitovirus 1 (Supp. Figure 1a). GcMV1 complete sequence is 2507 nt long having a 244 nt long untranslated region (UTR) at the 5' and an 85 nt long UTR at the 3' end. It encodes a putative RNA-dependent RNA polymerase (RdRp) of 725 amino acids translated only using the mitochondrial genetic code which shows a conserved mitoviruses RdRp domain (pfam pfam05919). The closer protein to GcMV1 RdRp, when analyzing it through BLASTP, is the RdRp encoded by Sclerotinia sclerotiorum mitovirus 15 (Table 1, Supp. Figure 1a). The family *Mitoviridae,* which was recently shown to include members infecting plants and insects (Chiapello et al., 2021; Nerva et al., 2019; Nibert et al., 2018), has been expanded in the last years, including many new species divided in four genera: *Unuamitovirus, Duamitovirus, Triamitovirus,* and *Kvaramitovirus.* Furthermore, a new basal clade having a further conserved amino-terminal domain (absent in our mitovirus from *Gnomonopsis*) was suggested for a new genus consideration (Ezawa et al., 2023). Since the putative RdRp encoded by GcMV1 is sufficiently distant from the closer hit to require a new species accommodating it in the family *Mitoviridae* (Table 1), we performed maximum likelihood phylogenetic analysis to show which genera would contain GcMV1. Results (Supp. Figure 1b) showed that GcMV1, the first mitovirus detected in its host *G. castaneae,* is a new member of the genus *Unuamitovirus.* We therefore propose for this virus species the binomial name *Unamitovirus gnomoniopsis*.

Molecular characterization of Gnomoniopsis castaneae chrysovirus 1

One viral genome showing homology to viruses from the family *Chrysoviridae* was detected with the NGS analysis and called Gnomoniopsis castaneae chrysovirus 1. The viral genome of GcCV1 has 4 genomic segments (named RNA1 to 4); each segment encodes one putative protein with the exception of RNA3 which encodes two putative proteins (Figure 1a). RACE analysis was performed to complete the sequences of the four genomic segments of GcCV1, allowing the identification of conserved sequences on both 5' and 3' of the genomic segment's ends (Figure 1b). Finally, the RdRp sequences of viruses belonging to the family *Chrysoviridae* were selected and aligned with the RdRp from GcCV1 (encoded by ORFA) to show, in (Figure 1c), that GcCV1 is likely a new member of the genus *Alphachrysovirus,* having as its closer relative the RdRp of Cryphonectria nitschkei chrysovirus 1 (CnCV1).

The first hit when blasting each putative protein against the NCBI nr database are shown in Table 1: as shown by the BLAST results, the putative protein encoded by ORFA, ORFB, ORFC and ORFE are commonly found in other member of the family *Chrysoviridae.* Indeed, ORFA encodes the putative RdRp, ORFB encodes the putative coat protein, ORFC encodes a replication associated protein (Liu et al., 2012) and ORFE encodes a cysteine protease. Interestingly, the first result when blasting ORFD-encoded protein is a putative nucleocapsid encoded by a virus called Plasmopara viticola lesion associated mymonavirus 1 (PvlaMiV1), related to the order *Mononegavirales.* In the case of CnCV1 (the closer virus to GcCV1 among the sequences contained in the nr database), small ORFs were reported for all the four segments composing the viral genome, and homology against the nucleocapsid from PvlaMiV1 was reported for the small ORF from RNA3 (called sORF3) as observed for GcCV1 (Shahi et al., 2021). Since the putative protein from CnCV1 RNA3 is not annotated on the sequence deposited on the NCBI database, it was impossible for us to detect the expected homology between GcCV1 ORFD and CnCV1 sORF3 using BLASTP. Thus, we submitted the protein encoded by GcCV1 ORFD to a TBLASTN search against the nt

database to detect hits against conserved protein that were not annotated on published chrysoviruses. Results in Supp. Table 4 show that the protein encoded by ORFD is conserved among a small number of chrysoviruses and other *Mononegavirales* related viruses.

To better describe conservation among the ORFD-like proteins and their distribution in the viruses related to the order *Mononegavirales*, we retrieved all the sequences from the results shown in Supp. Table 4. We then showed schematic representation of the genomic sequences retrieved from Table 2, highlighting the ORFD-like ORFs on each virus genomes to better show the variability between their position in all the genomes of the selected viruses (Figure 2). Interestingly, the chrysoviruses coding for the ORFD-like protein that are characterized in literature belong to a specific sub-clade within the genus *Alphachrysovirus*, as shown by phylogenetic analysis (Figure 1c).

After removing redundant and partial sequences, the RdRp encoding sequences from Table 2 were included in the list of viruses used by Pagnoni (Pagnoni et al. 2023) for building the phylogenetic tree including viruses from the order *Mononegavirales*. In the resulting tree (Figure 3), the ORFD-like encoding viruses were highlighted in bold to show that these proteins could be found in viruses belonging to different genera in the family *Mymonaviridae* e.g. the genus *Phyllomonavirus, Penicillimonavirus* and viruses found in one of the two clusters of the genus *Sclerotimonavirus.*

The level of identity among the identified ORFD-like proteins was investigated through protein alignment (Supp. Figure 2) and using the alignment to build an identity matrix including the identified proteins from chrysoviruses and *Mononegavirales*-related viruses. The list of protein sequences was cleaned from redundant sequences before aligning, and the results are shown in Supp. Figure 3. Overall, the identity observed among the ORFD-like proteins is quite low, and it is not possible to distinguish precisely the groups sharing highest homology using the genera they belong to as main criterion. Six residues were conserved among all the aligned proteins: using as a reference ORFD putative protein from GcCV1, the conserved amino acids were glycine (position 96), alanine (113), leucine (125), tryptophan (138), tyrosine (141) and asparagine (189) (Supp. Figure 2).

Given the low identity among ORFD-like proteins, we performed structure prediction using AlphaFold2 to compare the predicted structural conservation. We selected complete ORFD-like proteins from the chrysoviruses identified through BLAST analysis (Supp. Table 4): GcCV1, CnCV1, Verticillium dahliae chrysovirus 1 (VdCV1), and from two mymonaviruses: the closer sequence to GcCV1 ORFD (Plasmopara viticola lesion associated mymonavirus 1, PvlaMV1, and the more distant sequence Botrytis cinerea mymonavirus 1, BcMV1). Almost all the resulting structural models showed good values of predicted local distance difference test (pLDDT): the value obtained were among 73.6 and 87.3, which are usually connected to a good folding prediction (Tunyasuvunakool et al. 2021), with the only exception of BcMV1 ORFD-like that gave a pLDDT value of 38.6. Results in Supp. Figure 4a show that among chrysoviruses and PvlaMV1 it is possible to identify a large portion of the folded protein that is conserved with six alpha helixes and a beta sheet motif identified in the folded prediction (Supp. Figure 4b), while only a small motif could be detected when comparing GcCV1 ORFD to BcMV1. This is consistent with the results obtained from identity matrix (Supp. Figure 3) and from the protein sequence alignment (Supp. Figure 2) which show that BcMV1 ORFD-like sequence seem to diverge from all the other sequences identified.

Screening of the larger fungal collection

The presence of GcMV1 and GcCV1 was tested through real time RT-PCR in a collection of 102 further isolates of *G. castaneae* including isolates from different countries representing different continents (see Supp. Table 1 and Supp. Figure 5) to have a better perspective on the distribution of these viruses and possibly to observe possible mycovirus-induced phenotypes comparing a larger number of infected and non-infected isolates. The geographical origin of the fungal isolates is rather scattered across Europe (Supp. Table 1), with the majority of the fungal isolates collected in the Northern Italian regions of Valle d'Aosta, Friuli Venezia – Giulia, and Piedmont. Isolates from other Italian regions were also included in the further screening, such as Calabria, Campania and Lazio. A number of fungal isolates from neighboring Switzerland was also screened. The fungal collection also

included isolates from Spain, France, Czech Republic and Chile. Results on the detection of GcMV1 and GcCV1 are shown in Supp. Table 1, also including isolates analyzed through NGS. Overall, thirty-one fungal isolates were found infected with GcMV1 and forty-six isolates were found infected with GcCV1. In seven cases the two viruses were found co-infecting the same host. Even if the number of isolates from countries different from Italy was generally low, it is worth noting that GcCV1 was detected in Czech Republic and Switzerland, while GcMV1 was found only in Italian and Swiss isolates. Similar infection pattern was observed for Swiss and Italian isolates, as most of the Italian isolates originated from the Alpine-Northern area, not distant from Switzerland. Both viruses were detected in isolates from Chile, but only accumulated minimally.

Discussion

Since the relatively new detection of *G. castaneae* as cause of chestnut brown rot and, in some cases, bark cankers, few steps were made in the understanding of the mechanisms bringing to a change in the behavior of the fungus from an endophytic to a pathogenic lifestyle (Dobry et al., 2023). To date, different studies tried to characterize the putative interaction between *G. castaneae* and the Asian chestnut gall wasp (ACGW) as a factor boosting the impact of the fungal pathogen (Lione et al., 2016); furthermore, possible interactions with antagonistic fungi and bacteria such as *Trichoderma atroviride* and *Bacillus amyloliquefaciens* were tested for preventing chestnut brown rot (Pasche et al., 2016), but field studies did not support the positive results obtained in laboratory conditions and showed that the presence of both *G. castaneae* and *T. atroviride* is frequent in nature in *G. castaneae* affected crops (Meyer et al., 2015; Muñoz-Adalia et al., 2019). No studies, to date, took into consideration the possible interaction between *G. castaneae* and infecting viruses as a factor that could affect the virulence of the fungus on chestnut. This work provides two different viral models (a new member of the genus *Mitovirus,* GcMV1, and a new member of the genus *Alphachrysovirus*, GcCV1) to study possible mycovirus effects on the fungal host in the context of the endophytic/pathogenic interaction with chestnut trees, even if molecular tools to manipulate the fungus in order to obtain isogenic isolates infected and non-infected with the viruses are still to be developed.

G. *castaneae* is often found associated with chestnut trees as an endophyte, while it is not clear which condition is required to induce the pathogenic behavior. For this reason, it is hard to understand the origin of the fungus, if it is an exotic introduced fungus or if it has long been present as endophyte but became emergent because of changing environmental conditions. Our screening in a large collection of *G. castaneae* isolates showed that the identified viruses could be found easily in the northern Italian and in Swiss areas, which are close environments that could be subjected to natural or human exchanges in the fungal populations. However, the detection of the same viruses in isolates collected in the Czech Republic could be an indication of human spreading of such infected isolates. To our knowledge, the existence of distinct vegetative compatibility groups was never explored for *G. castaneae* populations, thus it is not clear how easily the mycoviruses could naturally spread among different fungal isolates. However, in principle, in order to explore the distribution and spread of *G. castaneae* in the world, viral diversity could be taken in consideration as a marker, together with phylogenetic studies on the fungal host (Thapa et al., 2016).

One of the most interesting results provided in this work is related to the evidence of a possible HGT event between viruses of distant phylogenetic origins. Indeed, the putative protein encoded by ORFD of GcCV1 (a putative new member of the genus *Alphachrysovirus*, Phylum *Duplornaviricota*) shows structural and sequence homology with putative proteins from viruses belonging to the family *Mymonaviridae* (Phylum *Negarnaviricota*). HGT events between viruses and hosts is supposed to represent a major factor in facilitating the evolutionary change in eukaryotes (Irwin et al., 2022); this phenomenon is easily observed in DNA viruses and retroviruses, but less frequent when considering HGT events between hosts and RNA viruses due to the lack of DNA phase of the latter (Taylor et al., 2009). HTG among RNA viruses was already hypothesized in different cases: interestingly, many of these studies focused on mycoviral models. Indeed, Liu and coworkers investigated the putative cross

family HGT events between dsRNA viruses as a factor increasing viral diversity (Liu et al., 2012), and an HGT event regarding a nidovirus-related helicase was supposed for a hypovirus found in *Rhizoctonia solani* (Abdoulaye et al., 2021). The evidence collected on distant HTG events among RNA viruses infecting fungi could be related to the frequent virus co-infection observed for these hosts (confirmed also in our *G. castaneae* collection), allowing the replication of multiple unrelated infectious agents in the same cellular environment (Picarelli et al., 2019), also having a role in the cross kingdom spread of viruses from plant to fungi (Andika et al., 2023). In the case of the chrysoviruses, it has been shown that viral segments could encode small putative proteins at the 5'; in the majority of the cases, the small ORFs encoded for protein showing no conservation against the public databases (Wu et al., 2023; Shah et al., 2023). Nevertheless, the data collected on the ORFD-encoded putative protein of GcCV1 suggest HGT events that likely led to the acquisition of the gene by a specific clade of chrysoviruses (Figure 1c) from mymonaviruses. This could be suggested by the fact that ORFD-like proteins are rarely found in phylogenetically close chrysoviruses, while homologs are easily found in different viruses belonging to the family *Mymonaviridae.* Sequence identity among different ORFD like proteins shows low level of conservation, and the same result is observed when superimposing the putative folded structure, with only a small motif resulting from the comparison between GcCV1 ORFD and the more distant homolog found through BLAST analysis (the ORFD-like sequence from Botrytis cinerea mymonavirus 1) allowing us to hypothesize an example of re-purposing of a virus gene, a common mechanism in virus evolution (Koonin et al., 2022). Nevertheless, the comparison between ORFD-like proteins from different chrysovirus and including the closest sequence belonging to a mymonavirus (Plasmopara viticola lesion associated mymonavirus 1) shows that a large portion of the folded structure is in this case conserved. No conserved domains could be detected in the ORFD-like putative proteins; only in few cases, these putative proteins were annotated as nucleocapsid (Chiapello et al., 2020) but a recent work unveiled the existence of RNA2 associated with these viruses belonging to the genus *Penicillimonavirus*, showing that their nucleocapsid is encoded by an ORF present on the second genomic fragment and the original annotation as nucleocapsid is likely wrong (Pagnoni et al., 2023). This evidence rise questions on the function of this group of homologous viral proteins both in negative sense RNA viruses and in chrysoviruses.

Conclusions

This work represents a first step in setting up a model for the study of fungal-virus systems in the context of a complex tripartite interaction between chestnut plants, the endophytic/pathogenic fungus *G. castaneae* and the viruses infecting the latter. Future studies would be focused on the possible effects of the mycoviruses GcMV1 and GcCV1 on *G. castaneae* and on the characterization of the function of the protein encoded by GcCV1 ORFD, for which no data on the possible activity is available to date.

Data availability

Reads used in this work are deposited in the NCBI (Bioproject: PRJNA1061088, Biosample: SAMN39254945, SRR27460082). The NCBI accession numbers for each viral segment identified in this work are found in Table 1.

Author contributions

Franco Ferilli: conceptualization, methodology, validation, writing. **Guglielmo Lione:** conceptualization, formal analysis, writing. **Paolo Gonthier**: conceptualization, formal analysis, writing. **Massimo Turina**: conceptualization, formal analysis, writing. **Marco Forgia**: conceptualization, formal analysis, methodology, experimental validation, writing.

Conflict of interest

Authors declare no conflict of interest.

Acknowledgements

This research was co-funded by Regione Piemonte through the F.E.A.S.R. 2014/2020, Projects #castagnopiemonte and 3C (Progetti pilota per la Cooperazione ed il miglioramento della Competitività della Castanicoltura regionale), through the activity of the Chestnut R&D Center, and by the European Commission through the programme INTERREG V-A Italy-Switzerland 2014/2020, Project MONGEFITOFOR id 540693 (linee guida per il MOnitoraggio e la Gestione delle Emergenze FITOsanitarie nelle FOReste delle Alpi centro-occidentali). Authors are grateful to Pierluigi Cavargna, Marielaure De La Harpe, Libor Jankovský, Mauro Jermini, Pasquale Nisi, Antonio Scalise and Vincenzo Tagliavento for providing fresh plant tissues or fungal isolates.

Bibliography

Abdoulaye, A. H., Hai, D., Tang, Q., Jiang, D., Fu, Y., Cheng, J., Lin, Y., Li, B., Kotta-Loizou, I., Xie, J., 2021. Two distant helicases in one mycovirus: evidence of horizontal gene transfer between mycoviruses, coronaviruses and other nidoviruses. *Virus Evolution, 7*(1). doi:10.1093/ve/veab043

Andika, I. B., Tian, M., Bian, R., Cao, X., Luo, M., Kondo, H., Sun, L., 2023. Cross-Kingdom Interactions Between Plant and Fungal Viruses. *Annual Review of Virology, 10*(1), 119-138. doi:10.1146/annurev-virology-111821-122539

Carugo, O., Pongor, S., 2001. A normalized root-mean-spuare distance for comparing protein three-dimensional structures. *Protein Science*, 10(7), 1470–1473.

Chiapello, M., Bosco, L., Ciuffo, M., Ottati, S., Salem, N., Rosa, C., Tavella, L., Turina, M., 2021. Complexity and Local Specificity of the Virome Associated with Tospovirus-Transmitting Thrips Species. *Journal of Virology, 95*(21), 10.1128/jvi.00597-00521

Choi, G. H., Nuss, D. L., 1992. Hypovirulence of chestnut blight fungus conferred by an infectious viral cDNA. *Science, 257*(5071), 800-803.

Chun, J., Yang, H. E., Kim, D. H., 2018. Identification of a Novel Partitivirus of *Trichoderma harzianum* NFCF319 and Evidence for the Related Antifungal Activity. *Frontiers in Plant Science, 9*. doi:10.3389/fpls.2018.01699

Deakin, G., Dobbs, E., Bennett, J. M., Jones, I. M., Grogan, H. M., Burton, K. S., 2017. Multiple viral infections in *Agaricus bisporu*s - Characterisation of 18 unique RNA viruses and 8 ORFans identified by deep sequencing. *Scientific Reports, 7*(1), 2469. doi:10.1038/s41598-017-01592-9

Dobry, E., Campbell, M., 2023. *Gnomoniopsis castaneae*: An emerging plant pathogen and global threat to chestnut systems. *Plant Pathology, 72*(2), 218-231. doi:https://doi.org/10.1111/ppa.13670

Espino-Vázquez, A. N., Bermúdez-Barrientos, J. R., Cabrera-Rangel, J. F., Córdova-López, G., Cardoso-Martínez, F., Martínez-Vázquez, A., Camarena-Pozos D. A., Mondo, S. J., Pawlowska, T. E., Abreu-Goodger, C., Partida-Martínez, L. P., 2020. Narnaviruses: novel players in fungal–bacterial symbioses. *The ISME Journal, 14*(7), 1743-1754. doi:10.1038/s41396-020-0638-y

Ezawa, T., Silvestri, A., Maruyama, H., Tawaraya, K., Suzuki, M., Duan, Y., Turina, M., Lanfranco, L., 2023. Structurally distinct mitoviruses: are they an ancestral lineage of the *Mitoviridae* exclusive to arbuscular mycorrhizal fungi (Glomeromycotina)? *mBio, 14*(4), e00240-00223. doi:10.1128/mbio.00240-23

Forgia, M., Chiapello, M., Daghino, S., Pacifico, D., Crucitti, D., Oliva, D., Ayllon, M. A., Turina, M., 2022. Three new clades of putative viral RNA-dependent RNA polymerases with rare or unique catalytic triads discovered in libraries of ORFans from powdery mildews and the yeast of oenological interest *Starmerella bacillaris*. *Virus Evolution, 8*(1). doi:10.1093/ve/veac038

Ghabrial, S. A., Castón, J. R., Jiang, D., Nibert, M. L., Suzuki, N., 2015. 50-plus years of fungal viruses. *Virology, 479*, 356-368.

Hollings, M., 1962. Viruses Associated with A Die-Back Disease of Cultivated Mushroom. *Nature*, 196, 962–965. Doi: https://doi.org/10.1038/196962a0

Irwin, N. A. T., Pittis, A. A., Richards, T. A., Keeling, P. J., 2022. Systematic evaluation of horizontal gene transfer between eukaryotes and viruses. *Nature Microbiology, 7*(2), 327-336. doi:10.1038/s41564-021-01026-3

Jumper, J., Evans, R., Pritzel, A., Green, T., Figurnov, M., Ronneberger, O., Tunyasuvunakool, K., Bates, R., Žídek, A., Potapenko, A., Bridgland, A., Meye, C., Kohl., S. A. A., Ballard, A. J., Cowie, A., Romera-Paredes, B., Nikolov, S., Jain, R., Adler, J., Back, T., Petersen, S., Reiman, D., Clancy, E., Zielinski, M., Steinegger, M., Pacholska, M., Berghammer, T., Bodenstein, S., Silver, D., Vinyals, O., Senior, A., W., Kavukcuoglu, K., Kohli, P., Hassabis, D., 2021. Highly accurate protein structure prediction with AlphaFold. *Nature, 596*(7873), 583-589.

Katoh, K., Rozewicki, J., Yamada, K. D., 2019. MAFFT online service: multiple sequence alignment, interactive sequence choice and visualization. *Briefings in bioinformatics, 20*(4), 1160-1166.

Kondo, H., Botella, L., Suzuki, N., 2022. Mycovirus Diversity and Evolution Revealed/Inferred from Recent Studies. *Annual Review of Phytopathology, 60*(1), 307-336. doi:10.1146/annurev-phyto-021621-122122

Koonin, E. V., Dolja, V., Krupovic, M., 2022. The logic of virus evolution. *Cell,* 30(7), 917-929.

Kotta-Loizou, I., 2021. Mycoviruses and their role in fungal pathogenesis. *Current Opinion in Microbiology, 63*, 10-18. doi:https://doi.org/10.1016/j.mib.2021.05.007

Langmead, B., Salzberg, S. L., 2012. Fast gapped-read alignment with Bowtie 2. *Nature methods, 9*(4), 357-359.

Lema, F., Baptista, P., Oliveira, C., & Ramalhosa, E., 2023. Brown Rot Caused by *Gnomoniopsis smithogilvyi* (syn. *Gnomoniopsis castaneae*) at the Level of the Chestnut Tree (*Castanea sativa* Mill.). *Applied Sciences, 13*(6), 3969.

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R., 2009. The sequence alignment/map format and SAMtools. *bioinformatics, 25*(16), 2078-2079.

Lione G., Giordano L., Sillo F., Gonthier. P., 2015. Testing and modelling the effects of climate on the incidence of the emergent nut rot agent of chestnut *Gnomoniopsis castanea*. *Plant Pathology* 64, 852-863

Lione, G., Danti, R., Fernandez-Conradi, P., Ferreira-Cardoso, J., Lefort, F., Marques, G., Meyer, J. B., Prospero, S., Radócz, L., Robin, C., Turchetti, T., Vettraino, A. M., Gonthier, P., 2019. The emerging pathogen of chestnut *Gnomoniopsis castaneae*: The challenge posed by a versatile fungus. *European Journal of Plant Pathology, 153*, 671-685.

Lione, G., Giordano, L., Ferracini, C., Alma, A., Gonthier, P., 2016. Testing ecological interactions between *Gnomoniopsis castaneae* and *Dryocosmus kuriphilus*. *Acta Oecologica, 77*, 10-17.

Liu, C., Li, M., Redda, E. T., Mei, J., Zhang, J., Wu, B., Jiang, X., 2019. A novel double-stranded RNA mycovirus isolated from *Trichoderma harzianum*. *Virology Journal, 16*(1), 113. doi:10.1186/s12985-019-1213-x

Liu, H., Fu, Y., Xie, J., Cheng, J., Ghabrial, S. A., Li, G., Peng, Y., Yi, Y., Jiang, D., 2012. Evolutionary genomics of mycovirus-related dsRNA viruses reveals cross-family horizontal gene transfer and evolution of diverse viral lineages. *BMC evolutionary biology, 12*, 1-15.

Meyer, J. B., Gallien, L., Prospero, S., 2015. Interaction between two invasive organisms on the European chestnut: does the chestnut blight fungus benefit from the presence of the gall wasp? *FEMS Microbiology Ecology, 91*(11), fiv122.

Milne, I., Stephen, G., Bayer, M., Cock, P. J., Pritchard, L., Cardle, L., Shaw, P. D., Marshall, D., 2013. Using Tablet for visual exploration of second-generation sequencing data. *Briefings in bioinformatics, 14*(2), 193-202.

Mirdita, M., Schütze, K., Moriwaki, Y., Heo, L., Ovchinnikov, S., Steinegger, M., 2022. ColabFold: making protein folding accessible to all. *Nature methods, 19*(6), 679-682.

Muñoz-Adalia, E. J., Rodríguez, D., Casado, M., Diez, J., Fernández, M., 2019. Fungal community of necrotic and healthy galls in chestnut trees colonized by *Dryocosmus kuriphilus* (Hymenoptera, Cynipidae). *iForest-Biogeosciences and Forestry, 12*(4), 411.

Nerva, L., Vigani, G., Silvestre, D. D., Ciuffo, M., Forgia, M., Chitarra, W., Turina, M., 2019. Biological and Molecular Characterization of Chenopodium quinoa Mitovirus 1 Reveals a Distinct Small RNA Response Compared to Those of Cytoplasmic RNA Viruses. *Journal of Virology, 93*(7), 10.1128/jvi.01998-01918. doi:10.1128/jvi.01998-18

Nibert, M. L., Vong, M., Fugate, K. K., Debat, H. J., 2018. Evidence for contemporary plant mitoviruses. *Virology, 518*, 14-24. doi:https://doi.org/10.1016/j.virol.2018.02.005

Nuss, D. L., 2005. Hypovirulence: mycoviruses at the fungal–plant interface. *Nature Reviews Microbiology, 3*(8), 632-642.

Pagnoni, S., Oufensou, S., Balmas, V., Bulgari, D., Gobbi, E., Forgia, M., Migheli, Q., Turina, M., 2023. A collection of Trichoderma isolates from natural environments in Sardinia reveals a complex virome that includes negative-sense fungal viruses with unprecedented genome organizations. *Virus Evolution, 9*(2). doi:10.1093/ve/vead042

Pasche, S., Crovadore, J., Pelleteret, P., Jermini, M., Mauch-Mani, B., Oszako, T., Lefort, F., 2016. Biological control of the latent pathogen *Gnomoniopsis smithogylvyi* in European chestnut grafting scions using *Bacillus amyloliquefaciens* and *Trichoderma atroviride*. *Dendrobiology, 75*.

Pearson, M. N., Beever, R. E., Boine, B., Arthur, K., 2009. Mycoviruses of filamentous fungi and their relevance to plant pathology. *Molecular plant pathology, 10*(1), 115-128.

Pettersen, E. F., Goddard, T. D., Huang, C. C., Meng, E. C., Couch, G. S., Croll, T. I., Morris, J. H., Ferrin, T. E., 2021, UCSF ChimeraX: Structure visualization for researchers, educators, and developers. *Protein Science*, 30 (1), 70-82.

Picarelli, M. A. S. C., Forgia, M., Rivas, E. B., Nerva, L., Chiapello, M., Turina, M., Colariccio, A., 2019. Extreme Diversity of Mycoviruses Present in Isolates of *Rhizoctonia solani* AG2-2 LP From *Zoysia japonica* From Brazil. *Frontiers in Cellular and Infection Microbiology, 9*. doi:10.3389/fcimb.2019.00244

R Developement Core Team, 2010. R: A language and environment for statistical computing.

Shah, U. A., Daudu, J. O., Filippou, C., Tubby, K. V., Coutts, R. H. A., Kotta-Loizou, I., 2023. Identification and sequence determination of a new chrysovirus infecting the phytopathogenic fungus *Dothistroma septosporum*. *Archives of virology, 168*(5), 144. doi:10.1007/s00705-023-05768-9

Shahi, S., Chiba, S., Kondo, H., Suzuki, N., 2021. Cryphonectria nitschkei chrysovirus 1 with unique molecular features and a very narrow host range. *Virology, 554*, 55-65.

Sillo, F., Giordano, L., Zampieri, E., Lione, G., De Cesare, S., Gonthier, P., 2017. HRM analysis provides insights on the reproduction mode and the population structure of *Gnomoniopsis castaneae* in Europe. *Plant Pathology, 66*(2), 293-303.

Sutela, S., Forgia, M., Vainio, E. J., Chiapello, M., Daghino, S., Vallino, M., Martino, E., Girlanda, M., Perotto, S., Turina, M., 2020. The virome from a collection of endomycorrhizal fungi reveals new viral taxa with unprecedented genome organization. *Virus Evolution, 6*(2). doi:10.1093/ve/veaa076

Taylor, D. J., Bruenn, J., 2009. The evolution of novel fungal genes from non-retroviral RNA viruses. *BMC Biology, 7*(1), 88. doi:10.1186/1741-7007-7-88

Thapa, V., Turner, G. G., Hafenstein, S., Overton, B. E., Vanderwolf, K. J., Roossinck, M. J., 2016. Using a Novel Partitivirus in *Pseudogymnoascus destructans* to Understand the Epidemiology of White-Nose Syndrome. *PLOS Pathogens, 12*(12), e1006076. doi:10.1371/journal.ppat.1006076

Tian, B., Xie, J., Fu, Y., Cheng, J., Li, B., Chen, T., Zhao, Y., Gao, Z., Yang, P., Barbetti, M. J., Tyler, B. M., Jiang, D., 2020. A cosmopolitan fungal pathogen of dicots adopts an endophytic lifestyle on cereal crops and protects them from major fungal diseases. *The ISME Journal, 14*(12), 3120-3135. doi:10.1038/s41396-020-00744-6

Tunyasuvunakool, K., Adler, J., Wu, Z., Green, T., Zielinski, M., Žídek, A., Bridgland, A., Cowie, A., Meyer, C., Laydon, A., Velankar, S., Kleywegt, G. J., Bateman, A., Evans, R., Pritzel, A., Figurnov, M., Ronneberger, O., Bates, R., Kohl S. A. A., Potapenko, A., Ballard A. J., Romera-Paredes, B., Nikolov, S., Jain R., Clancy, E., Reiman, D., Petersen, S., Senior, A. W., Kavukcuoglu, K., Birney, E., Kohli P., Jumper J., Hassabis, D., 2021. Highly accurate protein structure prediction for the human proteome. *Nature*, 596, 590–596.

Visentin, I., Gentile, S., Valentino, D., Gonthier, P., Cardinale, F., 2012. *Gnomoniopsis castanea* sp. nov. (Gnomoniaceae, Diaporthales) as the causal agent of nut rot in sweet chestnut. *Journal of Plant Pathology*, 411-419.

Voth, P. D., Mairura, L., Lockhart, B. E., May, G., 2006. Phylogeography of *Ustilago maydis* virus H1 in the USA and Mexico. *Journal of general virology, 87*(11), 3433-3441.

Wagemans, J., Holtappels, D., Vainio, E., Rabiey, M., Marzachì, C., Herrero, S., Ravanbakhsh, M., Tebbe, C. C., Ogliastro, M., Ayllón, M. M., Turina, M., 2022. Going Viral: Virus-Based Biological Control Agents for Plant Protection. *Annual Review of Phytopathology, 60*(1), 21-42. doi:10.1146/annurev-phyto-021621-114208

Wickham, H. (2007). Reshaping data with the reshape package. *Journal of statistical software, 21*, 1-20.

Wu, Z., Tian, X., Liu, X., Zhou, J., Yu, W., Qi, X., Peng, J., Hsiang, T., Wang, Q., Wu, N., Jiang, Y., 2023. Complete genome sequence of a novel chrysovirus infecting *Aspergillus terreus*. *Archives of virology, 168*(8), 209. doi:10.1007/s00705-023-05839-x

Xie, J., Jiang, D., 2014. New insights into mycoviruses and exploration for the biological control of crop fungal diseases. *Annual Review of Phytopathology, 52*, 45-68.

Zhou, L., Li, X., Kotta-Loizou, I., Dong, K., Li, S., Ni, D., Hong, N., Wang, G., Xu, W., 2021. A mycovirus modulates the endophytic and pathogenic traits of a plant associated fungus. *The ISME Journal, 15*(7), 1893-1906. doi:10.1038/s41396-021-00892-3

Figures and Tables

Table 1. List of viruses studied in the work.

Figure 1. a) schematic representation of the genome of GcCV1. Black lines represent the genomic sequence and colored arrows represent the ORFs. b) alignments of the conserved sequences on the 5’

and 3' of GcCV1 genomic ends. c) phylogenetic analysis of the viruses belonging to the family *Chrysoviridae*, including all the viruses accepted in the family by the ICTV. Asterisk are used to mark viruses encoding ORD-like putative protein.

Figure 2. Genomes of ORFD-like encoding *Mononegavirales* and *Chrysoviridae.* In each representation, the RdRp protein is shown in blue, the ORFD-like protein in green and the chrysovirus replication associated protein is shown in yellow. Erysiphe necator associated negative-stranded RNA virus 17 and Magnaporthe oryzae mymonavirus 2 are most likely partial genomes.

Figure 3. Phylogenetic analysis of ORFD-like encoding *Mononegavirales.*

Supp. Figure 1. a) schematic representation of the genome of GcMV1. Black lines represent the genomic sequence and colored arrows represent the ORFs. b) phylogenetic analysis of the viruses belonging to the family *Mitoviridae*, including all the viruses accepted in the family by the ICTV.

Supp. Figure 2. Protein alignment of ORFD-like putative proteins.

Supp. Figure 3. Identity matrix of the ORFD-like protein retrieved from the BLAST analysis in Supplementary Table 4.

Supp. Figure 4. a) Pairwise comparison of ORFD-like folded structures. b) Portion of the folded sequence of GcCV1 ORFD putative protein conserved with PvlaMV1. Only the portion of the protein sharing structural conservation was shown.

Supp. Figure 5. Map of distribution of European isolates of *G. castaneae.*

Supp. Table 1 List of fungal isolates used for this work and detection of GcMV1 and GcCV1. .

Supp. Table 2 List of primers used in this work.

Supp. Table 3 List of viruses used for the phylogenetic analysis of the family *Mitoviridae.*

Supp. Table 4 TBLASTN results for GcCV1 ORFD putative protein against the NCBI nr database.