Università degli Studi di Torino

Tesi di Dottorato di Ricerca in Scienze Biologiche
e Biotecnologie Applicate

PhD Thesis in Biological Sciences
and Applied Biotechnologies

Expanding the biodiversity of the RNA virosphere through characterization of fungal viromes

Marco Forgia

Academic Tutor: Prof. Giovanna Cristina Varese
Non-academic Tutor: Dr. Massimo Turina

XXXIII Cycle: 2017 – 2020

Contents

1 Introduction

1.1 Viruses: origin and evolution

Viruses are intracellular parasites able to exploit host's resources to replicate themselves. They were discovered during the XIX century as pathogens of human and plants, but years of research identified them as entities infecting organisms from all domain of life and from all kind of environments (Güemes *et al.*, 2016; Prangishvili, 2013; Myers *et al.*, 2020). They are ubiquitous, and recent studies associated them not only to pathogenic effects on the host, but also to mutualistic interactions and in some cases having important roles in controlling some important environmental parameter (Roossinck, 2011). An example of this, is represented by the role of viruses in regulating oceanic microorganisms community (Suttle, 2007). In this case viruses are the responsible of the lysis of around 20% of the microbial population every day causing the release of carbon, part of which is subtracted from the carbon cycle through sedimentation on the ocean surface, having a specific effect on carbon cycle with possible secondary effects on climate changes. Moreover, the lysis of the microorganisms also releases metals and nitrogen that are easily absorbed by remaining microbial cells.

Viruses are composed by genetic material encoding a relatively low number of proteins (but in some cases they can encode for a number of genes similar to that of small bacteria genomes). The genetic material could exploit both DNA-based and RNA-based strategies during their replication cycle; viral proteins can be divided in two core functions: i) the replication module (all the protein involved in the replication of the genetic material belonging to the virus) and ii) the virion morphogenesis module (that is important for transmission from cell to cell) (Krupovic *et al.*, 2019). The proteins belonging to the replication module are related to five groups: the first one contains the RNA dependent RNA polymerases (RdRps); these are found in all RNA viruses and can therefore be used as a marker gene (te Velthuis, 2014). The second group contains the reverse transcriptase (RT) found in both RNA and DNA retroviruses. Homologs of these proteins could be found in several cellular organisms, group II introns and different retroelements (Xiong and Eickbush, 1990). The third group shows protein-primed family B DNA polymerases (B DNAP), found in some dsDNA viruses and showing homology to DNA polymerases responsible for genome replication of archaea and eukaryotes (Kazlauskas *et al.*, 2016). The fourth proteins are the Rolling circle replication endonucleases (RCRE) typical of circular ssDNA viruses that are also responsible for replication of rolling circle replicating plasmids in bacteria (Chandler *et al.*, 2013; Zhao *et al.*, 2019). The last group of protein belonging to this core are the hexameric superfamily 3 helicases (S3H). The S3H domain can be found fused with other viral proteins and can be present in several different viruses with RNA, ssDNA and dsDNA genomes (Iyer *et al.*, 2005).

Inside the virion morphogenesis module are found all the different known coat proteins (CP). In this case, for most of the viral coat protein is not easy to find homologs with proteins encoded by cellular organisms and high diversification is found between different viral taxa. Nevertheless, some structural conservation is observed in some coat proteins and at least eight different classes of major capsid proteins were described and compared by Krupovic and Koonin (Krupovic and Koonin, 2017). The first group of capsid proteins (CPs) contains the most frequent CPs conformation among known viruses, the single jellyroll (SJR). These proteins share conformation homology with cellular carbohydrate binding proteins diffused in all three domains of life; this original function could have been crucial for early viruses to enter hosts cells through binding membrane carbohydrates bringing the virus close to the target. The second larger conformational group is represented by the double jellyroll (DJR) CPs. These proteins probably derived from a gene duplication and a fusion event between two SJR repetition and are found in dsDNA viruses. A third class of CPs is characterized by the HK97-like fold. These proteins are also found in dsDNA viruses infecting bacteria and archaea but in this case it is difficult to find conformational conservation in cellular proteins. Indeed, the only class of cellular protein sharing HK97-like fold are encapsulins, proteins found in bacterial and archaea nanocompartments which are probably derived from a domestication event of a viral coat protein. The fourth class of structure, the chymotrypsin-like protease fold, is found in Alphavirus CPs, but similar conformation can be found in cellular proteins. These CPs are part of a polyprotein with an N-terminal domain interacting with viral RNA, and a C-terminal domain that can process the CP from the polyprotein acting then as a structural component in the capsid. These proteins share the conformation with other viral non-structural proteases from flaviviruses that are evolved from the HtrA-like cellular proteases. Class five is represented by the helical nucleocapsids, that seems to have evolved from cellular nucleases. Class six is composed by retroviral Gag polyprotein: these are generally processed to produce three proteins: the matrix protein (MA), the capsid (CA) and the nucleocapsid (NC). The structure of the three proteins shares homology with different kind of cellular proteins: MA protein has a domain that is similar to the N terminal domain of different integrases. CA proteins share homology with the SCAN domain found in various vertebrate transcription factors in its C terminal portion; lastly NC protein binds the viral genome through Zn-knuckle motifs that is also found in cellular proteins. The matrix protein from arenaviruses, grouped in class seven, shows a RING domain in its conformation, which is also found in cellular proteins like ubiquitin ligases and that it was likely adapted to function in the virion morphogenesis module for arenaviruses. The last group of protein analyzed by Krupovic and Koonin contains the matrix proteins of mononegaviruses. These proteins share conformation with cellular cyclophilins, involved in the protein folding and incorporated from the cellular environment in several unrelated virions from coronaviruses to reoviruses and poxviruses. In this case, authors hypothesized that the gene encoding for cyclophilins was

incorporated by the viral genome and mutated to become the actual matrix protein typical for mononegaviruses.

Comparison among viruses is complicated by the lack of universal genes conserved among all known viruses: it is therefore hard to reconstruct the origin and evolution of these biological entities. To explain how viruses emerged in the world, three hypotheses have been proposed. The first hypothesis is the "virus early": in this theory viruses are residues of replicons from the pre-cellular world. These acquired the ability to enter the emerging cellular entities exploiting their resources for the replication cycle and evolved together with the emerging living organisms (Koonin *et al.*, 2006). The second hypothesis is the "regression": this hypothesis states that viruses could evolve from an intracellular parasite at the very beginning of DNA based cellular life. The intracellular parasite reduced his organization losing every structure but the genome encoding few proteins for replication and virion formation, bringing to the emergence of viruses as we now know them. In this scenario, the intercellular parasite emerged with the first DNA based cellular organisms spreading viruses in all the domains of life that evolved later (Abrahão *et al.*, 2017). The third hypothesis is the "escaped gene" hypothesis. In this case, viruses are selfish genetic elements that escaped from the organism of origin keeping the ability to spread from cell to cell to complete the replication cycle (Forterre and Krupovic., 2012).

Recently, a new theory was proposed that mixes the primordial origin and the escape origin of viruses. Based on phylogenetic analysis on key replication protein and coat proteins, this theory states that the replication module of viral protein was originated early in the evolution of life. Replication proteins originated from pre-cellular life world as selfish genetic elements that probably kept the ability to replicate in cells. The virion morphogenesis core instead, emerged from cellular proteins coopted to bind the viral genome and assemble the particle (Krupovic *et al.*, 2019).

1.2 Virus classification and phylogenetic interactions

Viruses can exploit several different strategies for their genome replication based on both RNA or DNA. It is therefore possible to identify all kind of genomic organization from ssDNA to dsDNA, ssRNA (both positive sense or negative sense) and dsRNA. During the early seventies, a classification system called Baltimore classification was proposed to organize viruses on the basis of the replication and genomic strategy (Baltimore, 1971). At the time, Baltimore proposed six classes, then a seventh was added when RNA based reverse-transcribing viruses were identified:

- Class one hosting dsDNA viruses.
- Class two hosting ssDNA viruses, often with a circular conformation and replicating via rolling circle replication (RCR) strategy.
- Class three hosting dsRNA viruses.
- Class four hosting positive sense ssRNA viruses.
- Class five hosting negative sense ssRNA viruses.
- Cass six hosting RNA reverse transcribing viruses.
- Class seven hosting DNA reverse transcribing viruses.

In recent years, the study of viral phylogenetic interactions brought to a total rearrangement of the viral taxonomical classification (Koonin *et al.*, 2020). Thanks to the great number of new viruses described using next generation sequencing (NGS) approaches in a broad range of different organisms (Shi *et al.*, 2016), a large amount of data is now available to perform better phylogenetic comparison and to find answers to some of the taxonomical and evolutionary question that are still open. Moreover, refined homology-based and structural-based comparisons among conserved viral proteins brought to a new perspective on the evolution of viruses. If the Baltimore classification can still be useful to explain and describe the different replication strategies, it has no value for describing the evolution and phylogenetic interactions among viruses. Some of the most interesting perspective on the new viral megataxonomy are discussed below.

From the reorganization of viral taxonomy, it is now accepted the hypothesis that all viruses encoding an RdRp or a RT protein (characterized by the conserved palm domain) are monophyletic. Using RdRp or RT sequences, it is possible to build a phylogenetic tree containing all RNA viruses and RT encoding entities (retroviruses and group II introns) that are grouped in a realm called *Riboviria* (Wolf et al., 2018) (Figure 1). Interestingly, phylogenetic analysis also suggests that dsRNA viruses emerged from two different events and that negative sense ssRNA viruses evolved from one of the two dsRNA viruses' branches. RNA viruses' tree is characterized by five branches containing the five newly identified Phyla: the first branch hosts the *Lenarviricota* phylum. This phylum contains the families *Leviviridae* (hosting RNA viruses infecting bacteria) and the eukaryotes infecting viruses from families *Botourmiaviridae, Mitoviridae* and *Narnaviridae.* All viruses from

this Phylum have a positive sense ssRNA genome. The second branch it's the Phylum *Pisuviricota.* In this branch, picorna-like viruses are hosted and one of the two events that brought to the emergence of dsRNA viruses took place. Indeed, dsRNA viruses from families *Picobirnaviridae, Amalgaviridae* and *Partitiviridae* (showing a dsRNA organization) are here grouped in the Class *Duplopiviricetes.* Branch number three contains the Phylum *Kitrinoviricota,* in this group, only positive sense ssRNA viruses from the Alpha-like supergroup are hosted. Branch four contains all the remaining viruses showing dsRNA genomes; these viruses are grouped in the Phylum *Duplornaviricota* that contains families like *Reoviridae, Chrysoviridae* and *Megabirnaviridae.* Branch five contains all negative sense ssRNA viruses in the Phylum *Negarnaviricota,* this branch is paraphyletic with branch number four as negative sense RNA viruses originated from dsRNA viruses similar to the ones found in Phylum *Duplornaviricota*.

In comparison to RNA viruses, It is much harder to reconstruct the phylogenetic interactions for viruses belonging to the remaining Baltimore classes (class one and two), but in both cases data suggest a polyphyletic origin that is more complicated than what explained for RNA viruses and retroviruses. Circular ssDNA viruses probably originated from bacterial plasmids replicating via rolling circle replication strategy; from phylogenetic studies it seems that bacterial and eukaryotes ssDNA viruses originated from three different events (Kazlauskas *et al.*, 2016). The plasmids from which the eukaryote viruses originated, showed probably a fusion between the REP endonucleases typical of this replication strategy and an S3Hs helicase (that is not found for prokaryotes ssDNA viruses). Moreover, the emergence of eukaryotic ssDNA viruses seems to be related to two different events. Given the polyphyletic nature of ssDNA viruses, the recent taxonomy review organized them in two different Realm (*Monodnaviria* and *Varidnaviria*) (Figure 2). Finally, dsDNA viruses show a great diversity and lack of conserved genes that make impossible to reconstruct a phylogenetic relation and a taxonomic organization at high levels. It seems moreover, that this kind of viruses evolved from many different events in different moments in the evolution of life (Iranzo *et al.*, 2016).

From the data discussed in this chapter, it is clear that Baltimore classification cannot have a value when discussing evolution of viruses. Indeed, looking at the original seven classes, we can find a common origin for all viruses belonging to class three to seven, originating from the same common ancestor. Moreover, class three viruses (dsRNA viruses) includes two groups of viruses that evolved independently from different branches at different times.

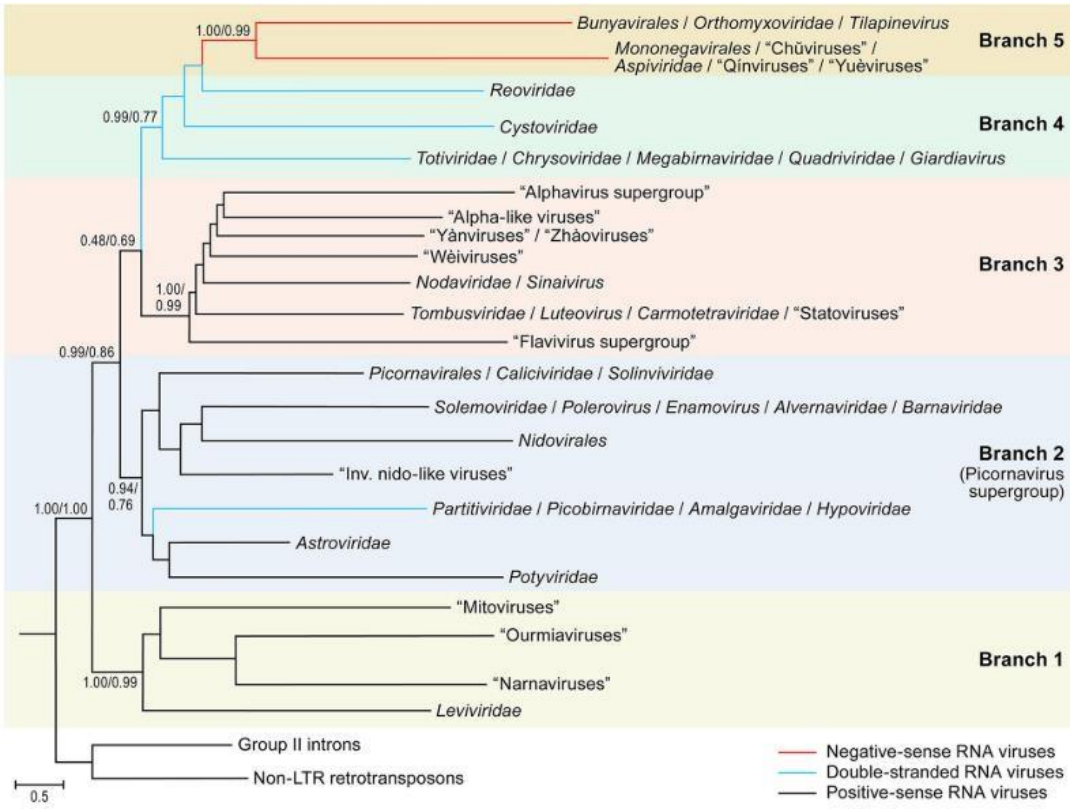


Figure 1. Phylogenetic analysis showing the five branches in the monophyletic tree obtained with RdRp data from RNA viruses. All the RNA viruses so fare described belong to the newly accepted Realm *Riboviria*. Colors of the branches described the genetic strategy for each viral taxon. (Wolf *et al.*, 2018).

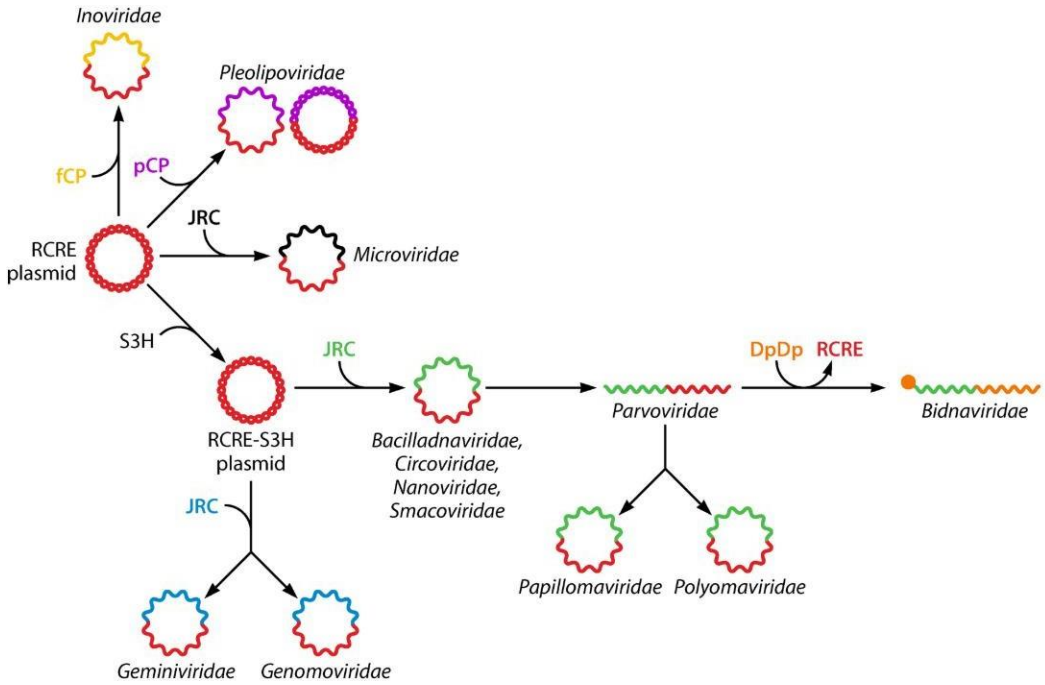


Figure 2. Origin of different lineage of circular ssDNA viruses. Different colors represent different origin of the genes encoded. *Inoviridae, Microviridae* and *Pleolipoviridae* (hosting viruses infecting bacteria and archaea) originated from rolling circle replication plasmids through fusion with different independent capsid proteins (fCP, filamentous capsid protein; pCP, polymorphic capsid protein; JRC, jelly roll capsid protein). Eukaryotes ssDNA viruses originated from Jelly roll capsid fusion with a S3H helicase-RCRE plasmid in different events: inone case giving rise to *Geminiviridae* and fungi infecting *Genomoviridae*, while the second event gave rise to different viral lineage. *Bidnaviridae* family contains linear ssDNA viruses that originated from substitution of a DNA-directed DNA polymerase with the REP protein following a reorganization of the genomic strategy. (Koonin *et al.*, 2020).

1.3 Mycoviruses

Mycoviruses (viruses that infects fungi) have been discovered in 1962 as pathogens of cultivated mushrooms inducing La France disease on *Agaricus bisporus* (Hollings*,* 1962) but interest in the field increased greatly after the discovery of the hypovirulence induced by the hypovirus Cryphonectria hypovirus 1 (CHV1) on the fungal host *Cryphonectria parasitica.* During the XX century, the ascomycetes pathogen *C. parasitica* was imported from Asia to America, then it reached Europe and started to kill the local population of chestnut inducing the disease called chestnut blight (Rigling and Prospero, 2018). American and European chestnut variety were severely affected by this disease and chestnut cultivation was endangered as affected plants were killed rapidly. In the 1950s, hypovirulent isolates started to emerge from *C. parasitica* infected chestnuts in Italy; these fungal strains were able to induce a mild lesion on chestnut without killing the trees and plants were able to coexist with the pathogen. Hypovirulent *C. parasitica* isolates started to spread naturally through Europe and scientists discovered that the change in hypovirulent *C. parasitica* ability to induce chestnut blight disease was caused by the infection of the hypovirus CHV1 (Milgroom and Cortesi, 2004; Dodds, 1980; Van Alfen, 1975). This first example of the potential as biocontrol agent brought to a growing interest in the mycoviral world and attempts to identify new mycoviruses were made with the aim to replicate what was observed for *C. parasitica* on other important fungal diseases. Initially, mycoviruses were searched through dsRNA extraction and cloning of the identified dsRNA profiles. This technique is time consuming and only few viruses could be identified through the years (specially dsRNA viruses and positive sense ssRNA viruses accumulating dsRNA at high level) (Nerva *et al.*, 2016); with the emergence of next generation sequencing approaches during the last years, the world of mycoviruses saw a tremendous expansion and mycoviruses could be identified infecting every branch of fungal phylogenetic tree (Bartholomaeus *et al.*, 2016; Neupane *et al.*, 2018; Turina *et al.*, 2018; Myers *et al.*, 2020; Nerva *et al.*, 2019a; Nerva *et al.*, 2019b; Chiapello *et al.*, 2020; Nerva *et al.*, 2016; Marzano *et al.*, 2016; Vainio *et al.*, 2015). Interestingly, almost all the mycoviruses identified are RNA viruses (both single strand positive or negative and double strand) with the only exception of three circular ssDNA viruses identified in recent years in *Sclerotinia sclerotiorum, Fusarium graminearum* and *Botrytis cinerea* (Yu *et al.*, 2010; Li *et al.*, 2020; Hao *et al.*, 2021)**.** RNA mycoviruses are spread in all five branches of the monophyletic RNA viral phylogenetic tree that was displayed in the paragraphs above, and they show some peculiar features compared to plant or animal viruses: mycoviruses doesn't have an extracellular phase (Ghabrial *et al.*, 2015); they are spread through isolates without escaping the cellular environment thanks to anastomosis. The only exception to this observation is represented by one of the ssDNA mycoviruses discovered (Liu *et al.*, 2016): in this case, researchers demonstrated that the virions could be passed from an infected isolated to a virus free one through insects feeding on the infected colony and then moved to the virus free one. This lack of extracellular phase can

also explain why a great number of identified mycoviruses don't produce any coat protein and don't protect the RNA genome in a virion. indeed, many mycoviruses have lost the ability to produce any protein except for the RdRp (for example fungal mitoviruses, narnaviruses and botourmiaviruses) and they just exist as naked RNAs even with a multipartite genome, as the case of Hadaka virus 1 (infecting *Fusarium oxysporum*) showing eleven segments not included in any virion structure (Sato *et al.*, 2020). Some of the most interesting features associated to mycoviruses have been described while characterizing ssDNA mycoviruses. The most peculiar ssDNA mycoviruses so far discovered were isolated from *Sclerotinia sclerotiorum* and *Fusarium graminearum* and named Sclerotinia sclerotiorum hypovirulence associated DNA virus 1 (SsHADV-1) and Fusarium graminearum gemytripvirus 1 (FgGMTV1), respectively (Yu *et al.*, 2010; Li *et al.*, 2020). SsHADV-1 was isolated from an hypovorulent *S. sclerotiorum* strain in 2010; it became an interesting model not only because until 2020 it has been the only case of a ssDNA mycovirus known or for the hypovirulent effect induced to their fungal host; indeed, further studies showed some of the most fascinating properties associated to a mycovirus. In facts, SsHADV-1 can infect virus-free fungal isolates just by spreading purified viral particles on the fungal colony (Yu *et al.*, 2013), and (first time for mycoviruses) viral transmission was observed when using a mycophagous insect as vector feeding on infected colonies and then moving it to virus-free colonies (Liu *et al.*, 2016). In 2019, authors working on this model system showed that SsHADV-1 infection cause *S. sclerotiorum* to change its lifestyle from pathogenetic to endophytic, and that plants colonized by endophytic *S. sclerotiorum* (SsHADV-1 infected) is protected from other fungal pests (Zhang *et al.*, 2020; Tian *et al.*, 2020). FgGMTV1, instead, was discovered and characterized from an hypovirulent isolate of *Fusarium graminearum* (Li *et al.*, 2020), and authors were able to build an infectious clone that allowed a fine characterization of the viral effect on the host. Compared to SsHADV-1 (which has a monopartite genome) FgGMTV1 has three genomic segments carrying the rep protein, the coat protein, and a third unknown protein that has a role in the control of viral infection. Experiments carried out through transfection of fungal protoplasts with the infectious clone brought to the isolation of fungal colonies lacking the third genomic segment; the comparison between these isolates and the normal FgGMTV1-infected strains showed that presence of the protein of unknown function encoded by the third ssDNA molecule induce higher viral DNA accumulation and milder viral effect on the host, balancing the pathogenic effect of FgGMTV1 on the fungus. As described, mycoviruses were first reported as pathogen, then their potential as hypovirulence inducer was discovered; however, most of the mycoviruses described recently don't seem to affect the phenotype of the infected fungus at least in laboratory condition. This apparently cryptic infection has been investigated through comparison of isogenic fungal isolates infected and non-infected with the mycoviruses, bringing to the identification of "hidden" phenotype changes induced by viral infections. The growing interest in this peculiar aspect of mycovirology gave birth to many studies unveiling potential viral roles in driving important aspects of

the fungal behavior and hypothesizing that these changes, even if hard to observe, could be important for fungal competition in the environment and could therefore explain why so many mycoviruses show this cryptic behavior when grown isolated in axenic conditions.

1.4 Hidden mycoviral phenotypes and potential roles in fungus–environment interaction

Identification of hidden phenotypes caused by mycoviruses on their hosts contributed to the idea that the phenotypic characterization of a fungal isolate need to take into consideration the fungus as a holobiont (focusing both on the fungus and on bacteria or viruses associated). Indeed, in some cases virus-induced effects have been correlated to positive effects on fungi of interest, showing a perspective on the viral role in evolution and adaptation that is really different from the common idea of viruses as pathogens. Even if evidences of hidden phenotypes changes are now reported frequently in literature, it is not always easy to correlate them to natural advantages; some example of mutualistic interaction between apparently hidden mycoviruses and hosts could be represented by the unclassified dsRNA virus Curvularia thermal tolerance virus (CThTV) that infects the fungus *Curvularia protuberata***,** an endophyte of the plant *Dichanthelium lanuginosum* (Marquez et al., 2007)**.** This plant lives in thermal soil over 50 C° in Yellowstone park (USA) and if not colonized by the fungal endophyte, it can not survive in this extreme condition; this ability is not observed even if the plant is colonized by a virus free fungal endophyte. Another well-known system is the yeast viral killer toxin in *Saccharomyces cerevisiae.* In this case, a totivirus is able to induce the production of a preprotoxin which is secreted and kills yeast cells not infected with the virus. The yeast totivirus, called ScV-L-A can show different satellites encoding for toxins, and despite the metabolic cost of the toxin production, it gives to the infected cell a great advantage in space colonization and resource availability (Schmitt and Breinig, 2006) to the point that the antiviral response (based in fungi on the DICER-AGO short RNAs pathway) is lost in *S. cerevisiae* probably to maintain the virus (Lax *et al.*, 2020). Other examples of mycovirus-induced effects on the host are represented by the increased osmotic tolerance in *C. parasitica* isolates transfected with mycoviruses isolated from a marine strain of *Penicillium aurantiogriseum* (Nerva et al., 2017) or the increased ochratoxin A production in *Aspergillus ochraceus* isolate infected with the partitivirus Aspergillus ochraceus virus (Nerva *et al.*, 2018a). In a recent work, the role of two narnaviruses was investigated in association of an endobacteria co-infecting *Rhizopus microspores*. Narnaviruses (named RmNV-20S and RmNV-23S) were found to influence sexual and asexual reproduction in the host, bringing to an increase in asexual sporulation when the viruses were absent compared to higher level of sexual reproduction in the presence of them (Espino-Vazquez *et al.*, 2020). In *Trichoderma* species, commonly used as biocontrol agent against fungal pathogens, the characterization of the mycovirus Trichoderma harzianum mycovirus 1 (ThMV1) showed that infection with ThMV1 reduced the biocontrol efficiency while promoting the growth of plants colonized by the infected fungus (Liu *et al.*, 2019). Moreover, studies on *Trichoderma harzianum* showed that the partitivirus ThPV1 induce an higher efficiency in the inhibition of two fungal pathogen (*Rhizoctonia solani* and *Pleurotus ostreatus*) correlated to higher expression of the gene encoding for β-1,3-

glucanase (Chun *et al.*, 2018). Lastly, a potential role of mycoviruses associated to arbuscular mycorrhizal fungi (AM fungi) in the production of small RNAs (sRNAs) through the antiviral response was hypothesized for mycoviruses associated to *Gigaspora margarita* (Silvestri *et al.*, 2020). These viral sRNAs could be detected in the AM fungi colonized roots of *Medicago truncatula* as sRNA exchange occurs in this interaction, potentially affecting plant RNA regulation through DICER-AGO pathway. In these cases, the observed phenotypes could be important for space colonization and competition in the natural environment even if the demonstration of an effect is not always obtained in natural condition as showed for CThTV**.** It is interesting to point out that mycoviruses reported in possible beneficial interaction for the host belong to many different taxonomical groups from double stranded to single stranded RNA viruses, so these abilities are more likely related to coevolution events between virus and host rather than to features typical from a precise viral taxonomical unit. Thus, the study of the virome associated to interesting fungi could be highly relevant to understand the cause of beneficial effects on fungi with a mutualistic behavior (for example mycorrhizal or endophytic fungi), to unveil potential pathogenic effect on interesting fungi (edibles or mutualistic) or even for their possible application as biocontrol agent.

1.5 ORFans sequences: the hidden information in virome investigations

As explained in the paragraphs above, the investigation of mycoviruses infecting collection of fungi could be relevant to characterize potential mutualistic or pathogenic interaction suitable for translational applications on fungi of interest. Moreover, viruses characterized through the HGT process could also bring to a deeper comprehension of the viral world, allowing for example the identification of new viral taxa (Chiapello *et al.*, 2020). Our knowledge on mycoviral diversity saw a great increase thanks to NGS approaches, but still a lot of new viral taxa are emerging from high throughput analysis suggesting how this field is still growing. In this framework particular interest is given to the sequences known as ORFans (Deakin *et al.*, 2017); these sequences are RNA molecules encoding for putative proteins of unknown function without a match on the NCBI non redundant database, and not originating from a DNA sequence integrated in the host genome or episomic DNA. Most of the bioinformatic pipelines used to mine new viruses from transcriptomic data use an homology-based approach to detect viral sequences, comparing the putative protein produced by the transcriptome contigs to a viral protein database (Nerva *et al.*, 2018b). This approach implies that undiscovered viruses showing low homology to known ones or viral genomic fragments never associated to a viral taxon have few chances to be detected using the standard pipeline. By characterizing the ORFans hidden in each of our virome analysis, we aim at detecting new and interesting viral entities that push further our perspective on the RNA virus world or that complete viral genomes that have been partially characterized so far as in the discovery of orphanplasmoviruses, a group of viruses encoding an RdRP that was not sufficiently conserved to be detected by homology based methods (Chiapello *et al.*, 2020).

1.6 Aim of this work

Our aim was to investigate the viral diversity characterizing collection of fungi that are relevant for their role in the environment (whether as pathogens or as mutualistic fungi), through meta-transcriptomic analysis. This approach gave us the opportunity to identify several new viral entities that represent a starting point in defining new viral taxa and that can not always easily fit inside the approved concept of the monophyletic evolution of RNA viruses. The taxonomical and phylogenetic implications unveiled by our results could in the future result in the identification of interactions between the virus and the host, and between the infected fungus and the colonized environment, possibly helping us to better understand the role and importance of the viral presence in the context of multitrophic interactions.

1.7 References

Abrahão J. S., Araújo R., Colson P., Scola B. L. (2017) The analysis of translation-related gene set boosts debates around origin and evolution of mimiviruses. *PLOS Genetics*.

Baltimore D. (1971) Expression of animal virus genomes. *Bacteriological reviews,* **35**, 235.

Bartholomaeus A., Wibberg D., Winkler A., Puehler A., Schlueter A., Varrelmann M. (2016) Deep Sequencing Analysis Reveals the Mycoviral Diversity of the Virome of an Avirulent Isolate of Rhizoctonia solani AG-2-2 IV. *Plos One,* **11**.

Chandler M., Cruz F. d. l., Dyda F., Hickman A. B., Moncalian G., Ton-Hoang B. (2013) Breaking and joining single-stranded DNA: the HUH endonuclease superfamily. *Nature reviews. Microbiology,* **11**, 525-538.

Chiapello M., Rodriguez-Romero J., Ayllon M. A., Turina M. (2020) Analysis of the virome associated to grapevine downy mildew lesions reveals new mycovirus lineages. *Virus evolution,* **6**, veaa058-veaa058.

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**.

Deakin G., Dobbs E., Bennett J. M., Jones I. M., Grogan H. M., Burton K. S. (2017) Multiple viral infections in Agaricus bisporus - Characterisation of 18 unique RNA viruses and 8 ORFans identified by deep sequencing. *Scientific Reports,* **7**.

Dodds J. A. (1980) Association of type 1 viral-like dsRNA with club-shaped particles in hypovirulent strains of Endothia parasitica. *Virology,* **107**, 1-12.

Espino-Vazquez A. N., Bermudez-Barrientos J. R., Cabrera-Rangel J. F., Cordova-Lopez G., Cardoso-Martinez F., Martinez-Vazquez A., Camarena-Pozos D. A., Mondo S. J., Pawlowska T. E., Abreu-Goodger C., Partida-Martinez L. P. (2020) Narnaviruses: novel players in fungal-bacterial symbioses. *Isme Journal*.

Forterre P., Krupovic. M. (2012) The Origin of Virions and Virocells: The Escape Hypothesis Revisited. *In: Viruses: Essential Agents of Life*, pp. 43-60. Springer.

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

Güemes A. G. C., Youle M., Cantú V. A., Felts B., Nulton J., Rohwer F. (2016) Viruses as Winners in the Game of Life. *Annual Review of Virology,* **3**, 197-214.

Hao F., Wu M., Li G. (2021) Characterization of a novel genomovirus in the phytopathogenic fungus Botrytis cinerea. *Virology,* **553**, 111-116.

Hollings M. (1962) Viruses associated with a die-back disease of cultivated mushroom. *Nature,* **196**, 962-965.

Iranzo J., Krupovic M., Koonin E. V., Hendrix R., Maslov S., Segall A. (2016) The Double-Stranded DNA Virosphere as a Modular Hierarchical Network of Gene Sharing. *mBio*.

Iyer L., Koonin E. V., Leipe D. D., Aravind L. (2005) Origin and evolution of the archaeo-eukaryotic primase superfamily and related palm-domain proteins: structural insights and new members. *Nucleic acids research,* **33**, 3875-3896.

Kazlauskas D., Krupovic M., Venclovas C. (2016) The logic of DNA replication in double-stranded DNA viruses: insights from global analysis of viral genomes. *Nucleic acids research,* **44**, 4551–4564.

Koonin E. V., Dolja V. V., Krupovic M., Varsani A., Wolf Y. I., Yutin N., Zerbini F. M., Kuhn J. H. (2020) Global Organization and Proposed Megataxonomy of the Virus World. *Microbiology and molecular biology reviews : MMBR,* **84**.

Koonin E. V., Senkevich T. G., Dolja V. V. (2006) The ancient Virus World and evolution of cells. *Biology Direct,* **1**, 29.

Krupovic M., Dolja V. V., Koonin E. V. (2019) Origin of viruses: primordial replicators recruiting capsids from hosts. *Nature Reviews Microbiology,* **17**, 449-458.

Krupovic M., Koonin E. V. (2017) Multiple origins of viral capsid proteins from cellular ancestors. *Proceedings of the National Academy of Sciences of the United States of America,* **114**, E2401-E2410.

Lax C., Tahiri G., Patiño-Medina J. A., Cánovas-Márquez J. T., Pérez-Ruiz J. A., Osorio-Concepción M., Navarro E., Calo S. (2020) The Evolutionary Significance of RNAi in the Fungal Kingdom. *International Journal of Molecular Sciences,* **21**, 9348.

Li P., Wang S., Zhang L., Qiu D., Zhou X., Guo L. (2020) A tripartite ssDNA mycovirus from a plant pathogenic fungus is infectious as cloned DNA and purified virions. *Science Advances,* **6**, 14.

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**.

Liu S., Xie J., Cheng J., Li B., Chen T., Fu Y., Li G., Wang M., Jin H., Wan H. (2016) Fungal DNA virus infects a mycophagous insect and utilizes it as a transmission vector. *Proceedings of the National Academy of Sciences,* **113**, 12803-12808.

Marquez L. M., Redman R. S., Rodriguez R. J., Roossinck M. J. (2007) A virus in a fungus in a plant: Three-way symbiosis required for thermal tolerance. *Science,* **315**, 513-515.

Marzano S.-Y. L., Nelson B. D., Ajayi-Oyetunde O., Bradley C. A., Hughes T. J., Hartman G. L., Eastburn D. M., Domier L. L. (2016) Identification of Diverse Mycoviruses through Metatranscriptomics Characterization of the Viromes of Five Major Fungal Plant Pathogens. *Journal of Virology,* **90**, 6846-6863.

Milgroom M. G., Cortesi P. (2004) Biological control of chestnut blight with hypovirulence: a critical analysis. *Annu. Rev. Phytopathol.,* **42**, 311-338.

Myers J. M., Bonds A. E., Clemons R. A., Thapa N. A., Simmons D. R., Carter-House D., Ortanez J., Liu P., Miralles-Durán A., Desirò A., Longcore J. E., Bonito G., Stajich J. E., Spatafora J. W., Chang Y., Corrochano L. M., Gryganskyi A., Grigoriev I. V., James T. Y., Taylor J. W. (2020) Survey of Early-Diverging Lineages of Fungi Reveals Abundant and Diverse Mycoviruses. *mBio*.

Nerva L., Chitarra W., Siciliano I., Gaiotti F., Ciuffo M., Forgia M., Varese G., Turina M. (2018a) Mycoviruses mediate mycotoxin regulation in Aspergillus ochraceus. *Environmental microbiology*.

Nerva L., Ciuffo M., Vallino M., Margaria P., Varese G. C., Gnavi G., Turina M. (2016) Multiple approaches for the detection and characterization of viral and plasmid symbionts from a collection of marine fungi. *Virus Research,* **219**, 22-38.

Nerva L., Forgia M., Ciuffo M., Chitarra W., Chiapello M., Vallino M., Varese G. C., Turina M. (2019a) The mycovirome of a fungal collection from the sea cucumber Holothuria polii. *Virus research,* **273**, 197737-197737.

Nerva L., Silvestri A., Ciuffo M., Palmano S., Varese G., Turina M. (2017) Transmission of Penicillium aurantiogriseum partiti-like virus 1 to a new fungal host (Cryphonectria parasitica) confers higher resistance to salinity and reveals adaptive genomic changes. *Environmental Microbiology*.

Nerva L., Turina M., Zanzotto A., Gardiman M., Gaiotti F., Gambino G., Chitarra W. (2019b) Isolation, molecular characterization and virome analysis of culturable wood fungal endophytes in esca symptomatic and asymptomatic grapevine plants. *Environmental Microbiology,* **21**, 2886-2904.

Nerva L., Varese G. C., Turina M. (2018b) Different Approaches to Discover Mycovirus Associated to Marine Organisms. *Methods in molecular biology (Clifton, N.J.),* **1746**, 97-114.

Neupane A., Feng C. C., Feng J. H., Kafle A., Bucking H., Lee Marzano S. Y. (2018) Metatranscriptomic Analysis and In Silico Approach Identified Mycoviruses in the Arbuscular Mycorrhizal Fungus Rhizophagus spp. *Viruses-Basel,* **10**, 12.

Prangishvili D. (2013) The Wonderful World of Archaeal Viruses. *Annual Review of Microbiology,* **67**, 565-585.

Rigling D., Prospero S. (2018) Cryphonectria parasitica, the causal agent of chestnut blight: invasion history, population biology and disease control. *Molecular Plant Pathology,* **19**, 7-20.

Roossinck M. J. (2011) The good viruses: viral mutualistic symbioses. *Nature Reviews Microbiology,* **9**, 99-108.

Sato Y., Shamsi W., Jamal A., Bhatti M. F., Kondo H., Suzuki N. (2020) Hadaka Virus 1: a Capsidless Eleven-Segmented Positive-Sense Single-Stranded RNA Virus from a Phytopathogenic Fungus, Fusarium oxysporum. *mBio,* **11**.

Schmitt M. J., Breinig F. (2006) Yeast viral killer toxins: lethality and self-protection. *Nature Reviews Microbiology,* **4**, 212-221.

Shi M., Lin X.-D., Tian J.-H., Chen L.-J., Chen X., Li C.-X., Qin X.-C., Li J., Cao J.-P., Eden J.-S. (2016) Redefining the invertebrate RNA virosphere. *Nature,* **540**, 539-543.

Silvestri A., Turina M., Fiorilli V., Miozzi L., Venice F., Bonfante P., Lanfranco L. (2020) Different Genetic Sources Contribute to the Small RNA Population in the Arbuscular Mycorrhizal Fungus Gigaspora margarita. *Frontiers in Microbiology,* **11**.

Suttle C. A. (2007) Marine viruses - major players in the global ecosystem. *Nature Reviews Microbiology,* **5**, 801-812.

te Velthuis A. J. W. (2014) Common and unique features of viral RNA-dependent polymerases. *Cellular and Molecular Life Sciences,* **71**, 4403-4420.

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. *Isme Journal*.

Turina M., Ghignone S., Astolfi N., Silvestri A., Bonfante P., Lanfranco L. (2018) The virome of the arbuscular mycorrhizal fungus Gigaspora margarita reveals the first report of DNA fragments corresponding to replicating non-retroviral RNA viruses in Fungi. *Environmental microbiology*.

Vainio E. J., Jurvansuu J., Streng J., Rajamaki M. L., Hantula J., Valkonen J. P. T. (2015) Diagnosis and discovery of fungal viruses using deep sequencing of small RNAs. *Journal of General Virology,* **96**, 714-725.

Van Alfen N. K., Jynes, R.A., Anagnostakis, S.L., Day, P.R., (1975) Chestnut blight: biological control by transmissible hypovirlence in Endothia parasitica. *Science,* **189**, 890-891.

Wolf Y. I., Kazlauskas D., Iranzo J., Lucia-Sanz A., Kuhn J. H., Krupovic M., Dolja V. V., Koonin E. V. (2018) Origins and Evolution of the Global RNA Virome. *Mbio,* **9**.

Xiong Y., Eickbush T. (1990) Origin and evolution of retroelements based upon their reverse transcriptase sequences. *The EMBO journal,* **9**, 3353-3362.

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. *Proceedings of the National Academy of Sciences of the United States of America,* **107**, 8387-8392.

Yu X., Li B., Fu Y., Xie J., Cheng J., Ghabrial S. A., Li G., Yi X., Jiang D. (2013) Extracellular transmission of a DNA mycovirus and its use as a natural fungicide. *Proceedings of the National Academy of Sciences of the United States of America,* **110**, 1452-1457.

Zhang H., Xie J., Fu Y., Cheng J., Qu Z., Zhao Z., Cheng S., Chen T., Li B., Wang Q., Liu X., Tian B., Collinge D. B., Jiang D. (2020) A 2-kb Mycovirus Converts a Pathogenic Fungus into a Beneficial Endophyte for Brassica Protection and Yield Enhancement. *Molecular Plant,* **13**, 1420-1433.

Zhao L., Rosario K., Breitbart M., Duffy S. (2019) Eukaryotic circular rep-encoding single-stranded DNA (CRESS DNA) viruses: Ubiquitous viruses with small genomes and a diverse host range. *Adv Virus Res,* **103**, 71-133.

Chapter 2

Extreme Diversity of Mycoviruses Present in Isolates of *Rhizoctonia solani* AG2-2 LP From *Zoysia japonica* From Brazil

Maria Aurea S. C. Picarelli [1†], Marco Forgia[2,3†], Eliana B. Rivas [4], Luca Nerva[3,5], Marco Chiapello[3], Massimo Turina[3] and Addolorata Colariccio[1]

† First and second author equally contributed to the work

1 Plant Virology Laboratory, Instituto Biológico, São Paulo, Brazil,
2 Department of Life Science and System Biology, University of Turin, Turin, Italy,
3 Institute for Sustainable Plant Protection, CNR, Turin, Italy,
4 Phytopathological Diagnostic Laboratory, Instituto Biológico, São Paulo, Brazil,
5 Council for Agricultural Research and Economics—Research Centre for Viticulture and Enology CREA-VE, Conegliano, Italy.

Published on Fontiers in Cellular and Infection Microbiology
DOI: https://doi.org/10.3389/fcimb.2019.00244

2.1 Abstract

Zoysia japonica, in Brazil, is commonly infected by *Rhizoctonia solani* (*R. solani*) in humid and cool weather conditions. Eight isolates of *R. solani*, previously identified as belonging to the AG2-2 LP anastomosis group, isolated from samples from large path symptoms, were collected from three counties in São Paulo state (Brazil) and investigated for the presence of mycoviruses. After detection of double-strand RNA (dsRNA) in all samples, RNA_Seq analysis of ribosomal RNA-depleted total RNA from *in vitro* cultivated mycelia was performed. Forty-seven partial or complete viral unique RNA dependent-RNA polymerase (RdRp) sequences were obtained with a high prevalence of positive sense ssRNA viruses. Sequences were sufficiently different from the first match in BLAST searches suggesting that they all qualify as possible new viral species, except for one sequence showing an almost complete match with Rhizoctonia solani dsRNA virus 2, an alphapartitivirus. Surprisingly four large contigs of putative viral RNA could not be assigned to any existing clade of viruses present in the databases, but no DNA was detected corresponding to these fragments confirming their viral replicative nature. This is the first report on the occurrence of mycoviruses in *R. solani* AG2-2 LP in South America.

Keywords: mycoviruses, *Rhizoctonia solani*, grass, multiple infection, viral diversity, phylogenetic analysis, virus taxonomy

2.2 Introduction

Zoysia japonica (*Z. japonica*) Steud, especially from the cultivar 'Esmeralda', comprises 81% of the cultivated grasses in Brazil (Zanon, 2015). In 2015, the sod production of cultivated grasses reached 24,000 hectares (Antoniolli, 2015), which is an increase of 40% between 2010 and 2015 (Zanon, 2015). Nevertheless, it is very susceptible to the large patch disease caused by the fungus *Rhizoctonia solani* (*R. solani*) AG 2-2 LP, considered the most important disease of zoysia grass worldwide. Disease control is difficult, and practices employed in management are hardly effective, also, because fungicides are not approved on public areas or home gardens. Considering these obstacles, biological control agents are desirable alternatives for disease management thanks to their environmental safety.

Since the first report in the 1960s, mycoviruses were searched for and found in many classes of phytopathogenic fungi, mostly because some mycoviruses can reduce the capacity of fungi to cause disease and may have the potential application as biological control agents. More recently, mycoviruses were also shown to be important for their environmental role and for modulating intra and inter-species interactions (Nerva et al., 2018a; Nerva et al., 2017; Chun et al., 2018; Drinnenberg et al., 2011), possibly mediated by complex tripartite symbiotic relationships (Marquez et al., 2007). Not all the viruses associated with fungal pathogens affect virulence, but some can indeed cause

hypovirulence, as is the case of the classic model system Cryphonectria hypovirus 1, infecting *Cryphonectria parasitica* (Nuss, 2005; Turina and Rostagno, 2007), or the association between Rhizoctonia solani partitivirus 2 (RsPV2) and *R. solani* AG-1 IA causing hypovirulence on rice (Zheng et al., 2014).

The presence of double-stranded RNAs (dsRNAs) is evidence of a mycovirus infection, which have been reported in many phytopathogenic fungi species and in different anastomosis groups from various hosts of *R. solani* (Bharathan et al., 2005; Strauss et al., 2000; Das et al., 2016; Zheng et al., 2018). More recently, ssDNA viruses were also shown to infect phytopathogenic fungi (Yu et al., 2010).

Nowadays, most approaches used to characterise fungal viruses rely on Next-Generation Sequencing (NGS) of total RNA depleted of ribosomal RNA or sequencing of small RNA (Vainio et al., 2015; Marzano and Domier, 2016; Marzano et al., 2016; Donaire and Ayllon, 2017). We directly compared the two methods in previous work and found that NGS of total RNA provides a more complete characterisation of fungal associated viruses (Nerva et al., 2016).

Mycoviruses in single and mixed infections have been previously detected in *R. solani*: Rhizoctonia solani virus 717 (*Betapartitivirus*), Rhizoctonia solani RNA virus HN008, Rhizoctonia solani dsRNA virus 1, Rhizoctonia solani partitivirus 2 and Rhizoctonia solani flexivirus 1 (unclassified virus) (Zhong et al., 2015; Bartholomäus et al., 2017; Zheng et al., 2018). Moreover, recently, a plant virus—cucumber mosaic virus—was shown to accumulate and replicate in *R. solani* (Andika et al., 2017). Further NGS characterisation of *R. solani*-associated viromes were performed on samples from USA (Marzano et al., 2016) and from samples from Germany (Bartholomaeus et al., 2017).

In a preliminary screen, we detected the presence of dsRNAs in eight asymptomatic Brazilian *R. solani* AG2-2 LP isolates, a pathogen of *Z. japonica,* from three counties of São Paulo state (Picarelli et al., 2015). All the dsRNA electrophoretic patterns showed 3 to 6 bands with different sizes, all greater than 2 kbp, and a fragment greater than 8 kbp (Picarelli, 2015). These complex electrophoretic patterns could be due to the presence of segmented viral genomes, mixed infections or defective dsRNAs. In this study, we aimed to characterise the virome associated with *R. solani*, isolated from *Z. japonica* grass that were positive in a preliminary dsRNA screen, to gather the first information about the diversity and the spread of mycoviruses in *R. solani* in different Brazilian regions. Although none of the isolates under scrutiny were hypovirulent, such a library of mycoviruses could be the basis for a targeted virus-induced gene silencing (VIGS) approach.

2.3 Materials and Methods

2.3.1 Fungal isolates origin and growth conditions

Z. japonica sheaths, showing large patch symptoms, were collected in three municipalities of São Paulo State, Brazil: Cotia (isolates IBRS07, IBRS15, IBRS16, and IBRS19), São Paulo (isolates IBRS04, IBRS22, and IBRS23), and Ilhabela (isolate IBRS11) (Supplementary Figure 1, online). All *R. solani* samples were collected from diseased patches of zoysia grass lawns showing the same characteristics. The eight *R. solani* AG2-2 LP isolates were maintained on potato dextrose agar medium at 25 °C, for a 12 h photoperiod (Picarelli et al., 2015). Long-term conservation of the fungal isolates was obtained by growing the fungi on paper strips stored at -80 °C or lyophilised mycelia after growth on potato dextrose broth and stored at -20 °C.

2.3.2 RNA extraction, DNA extraction and cDNA synthesis

Total RNA was extracted from 0.1 g of lyophilised fungal mycelium using the Spectrum™ Plant Total RNA Kit (Sigma-Aldrich, Darmstadt, Germany), according to the manufacturer's instructions. Copy DNA (cDNA) synthesis was performed using the High-Capacity cDNA Reverse Transcription Kit (Thermo Fisher Scientific, Waltham, MA, USA) as described in the kit's manual. DNA extraction was performed by breaking 50 mg of lyophilised mycelia in a bead beater, using 0.5 mm diameter glass beads, in a 2 mL Eppendorf tube with 700 µL of phenol and 700 µL of 2x STE-2%SDS. After centrifugation, the supernatant was collected and washed twice with chloroform-isoamyl alcohol, 24:1. The supernatant was then precipitated with 2 volumes of 100% ethanol and 0.1 volumes of 3 M sodium acetate, pH 5.2. The pellet was resuspended in 50 µL of H_2O, quantified with a Nanodrop 2000 (Thermo Fisher Scientific, Waltham, USA)) and diluted to 10 ng/µL for PCR applications.

2.3.3 Library preparations and bioinformatic analysis

Ribosomal RNA depletion, library preparations and Illumina sequencing were performed by Macrogen (Seoul, Republic of Korea); the assembly and virus identification steps were performed as previously described (Nerva et al., 2018b), where we have specified the details of commands and scripts used for each bioinformatics analysis step. Briefly, reads from RNA-Seq were assembled *de novo* using Trinity version 2.3.2 (Haas et al., 2013). Trinity assembly was then BLASTed against a custom viral database (https://osf.io/c9x2p/) to identify contigs of viral origin. The number of reads covering the viral genomes was obtained by mapping the reads from each sequenced library on reference sequences with Burrows-Wheeler Aligner (BWA) and Samtools (Li and Durbin, 2009; Li et al., 2009). The mapping step was performed as explained in detail previously (Nerva et al., 2018b) with only one exception: the bwa mem algorithm was used for the alignment instead of bwa aln. Mapping results were displayed using Tablet (Milne et al., 2013). Viral contigs displaying incomplete open reading frames (ORFs) and cut reads on the 5' or 3' ends were analysed using MITObim (Hahn, Bachmann and Chevreux, 2013) to attempt extending the incomplete 5' and 3' ends. After one iteration, the eventually extended sequences were used as a query for a BLAST search against the trinity assembly

to find contigs overlapping the extended region. Identified contigs were assembled using the CAP3 sequence assembly program (Huang and Madan, 1999).

2.3.4 Quantitative RT-PCR analysis

Primers for qRT-PCR were designed using Primer 3 (Untergasser et al., 2012), with the amplicon size between 70 and 120 bp. To associate specific RNA samples to each specific contig, qRT-PCR analysis were performed using a CFX Connect™ Real-Time PCR Detection System (Biorad, Hercules, USA). The PCR reaction was performed in 10 µL using the iTaq™ Universal SYBR® Green Supermix (Biorad, Hercules, USA). A melting curve analysis was performed at the end of the qRT-PCR protocol to check for unspecific PCR products. All the oligonucleotides used in the qRT-PCR protocol are reported in Supplementary Table 1, online.

2.3.5 ORF prediction and phylogenetic analyses

ORF predictions were performed using the ORF finder tool from NCBI, and predictions were made selecting the "standard" genetic code for all viral contigs, except the one closely related to mitoviruses, generally hosted in the mitochondria. These contigs were analysed selecting the "Mould, protozoan and coelenterate mitochondrial" genetic code. The putative function of the predicted protein was established by BLAST analysis, looking at the function of the closest proteins in the NCBI database. The predicted protein sequences were analysed through a BLASTP search using the domain finder option to evaluate the presence of any conserved domain in the sequence (such as the viral polymerase GDD conserved domain). Phylogenetic analyses were performed by aligning the viral RNA dependent-RNA polymerase (RdRp) proteins with MUSCLE implemented in MEGA6 (Tamura et al., 2013). Alignments were exported in FASTA format and submitted to the IQ-TREE web server (Trifinopoulos et al., 2016) to produce Maximum likelihood phylogenetic trees (Lam-Tung et al., 2015). The best substitution model was estimated automatically by IQ-TREE with ModelFinder (Kalyaanamoorthy et al., 2017) and ultrafast bootstrap analysis (Diep Thi et al., 2018) in which 1000 replicates were performed. For each tree, each specific model is indicated in the figure legend.

2.3.6 Amplification and cloning of fragments from the viral genomes

To confirm the sequence of regions of interest in the assembled contigs, fragments from some of the viral genomes were amplified by designing PCR primers based on the *in silico* assembly and performing PCR reactions on cDNA produced as described above (Supplementary Table 2, online). Using electrophoresis, amplified bands were separated on an agarose gel. Cut bands were cleaned using Zymoclean Gel DNA Recovery Kit (Zymo Research, Irvine, USA), inserted in the pGEMT vector using pGEM®-T Easy Vector Systems (Promega, Madison, USA) and transformed in chemically competent *E. coli* DH5α (Mix & Go! E.coli Transformation kit, Zymo Research, Irvine, USA). Colonies that were positive for the insertion were selected for Sanger sequencing (Biofab S.r.l., Rome, Italy).

2.4 Results and Discussion

A single sequencing run of the eight pooled *R. solani* isolates under scrutiny produced 167,355,298 total reads deposited in the SRA archive linked to BioProject PRJNA524447. After the Trinity run, a total of 89,779 contigs were assembled. A BLAST search of a custom prepared viral database identified a total of 56 putative viral contigs (Table 1). Among those, 44 contained the typical conserved motifs of a viral RdRp, which is essential for the replication of RNA viruses and often displays three conserved amino acids (GDD) that are crucial for the catalytic activity. Two are partial viral genomes, where the RdRp domain is probably located in the missing part, and one is a contig where the RdRp domain cannot be detected by search engines, but the protein sequence shows a similarity with the typical GDD RdRp domain from proteins belonging to the family *Hypoviridae*. Therefore, we identified at least 47 distinct viruses (Table 1). There were five contigs encoding Coat Protein (CP) of bipartite viruses, and finally four putative viral contigs, encoding for proteins of unknown function (ORFans). We then checked the association of each contig with each of the eight isolates using qRT-PCR specific for each fragment; the results are displayed in Table 2. The number of virus contigs/isolate varies from isolate IBRS15 containing 6 viral contigs to isolate IBRS23 containing 30 viral contigs. A quantitative estimation of the abundance of each contig can also be inferred by the number of mapped reads on each segment (Table 1).

The retrieved contigs showing homology with RdRps fell in 6 from the 16 clades that accommodate the overall viral diversity of RNA viruses of invertebrates (Shi et al. 2016).

2.4.1 Narna-levi related sequences

From the RNAseq assembly, we identified 22 sequences encoding for proteins showing high similarity with viruses from the Narna-levi clade (Shi et al. 2016). All these sequences encode for one ORF producing the putative RdRp; however, at least four of them were not complete. Since the RdRp domain was still detectable in the partial sequences, we included them in our phylogenetic analysis (Figure 1 and Figure 2). The predicted protein alignment and phylogenetic analysis showed that 18 sequences appeared to be part of the genus *Mitovirus*, while the 4 remaining were strongly related to ourmia-like viruses, but not included in the currently recognised *Narnavirus* genus. Thus, sequences were named as Rhizoctonia solani mitovirus 21 to 38 and Rhizoctonia solani ourmia-like virus 2 to 5, since previous work had already identified Rhizoctonia solani mitoviruses and ourmia-like viruses (Lakshman et al., 1998; Marzano et al., 2016; Bartholomäus et al., 2017). The amino acid sequence identity against the first hit in a BLAST search ranged from 80% (in the case of RsMV22 and a dsRNA viral element discovered in the same species) to 25.89% (in the case of RsOLV5 and Agaricus bisporus virus 15). From the phylogenetic tree (Figure 1), it is possible to observe that the mitoviruses detected in *R. solani* gather in four distinct sub-clades. Furthermore, we confirmed the necessity to revise the overall taxonomy of Levi-Narna viruses: the putative order Narnavirales should be established with a number of families, including the *Narnaviridae* (with the current *Narnavirus* genus) and the proposed/putative Mitoviridae (the current *Mitovirus* genus). The proposed new family Mitoviridae should be subdivided into a number of genera, including plant and fungal mitoviruses (Nibert et al., 2018; Nerva et al., 2019). The fact

that most *R. solani* mitoviruses fall in the same three clades is probably due to the fact that mitoviruses are located and replicate in the host mitochondria and interspecific transmission does not easily occur in nature. Nevertheless, at least four clades mixed with viruses from ascomycetes and basidiomycetes occur, suggesting that some horizontal transfer can still occur. Further pairwise comparison among the distinct viral contigs from the Trinity assembly showed that RsMV 21 and 36 and RsMV 28 and 30 were almost completely identical (98% and 100% at the nucleotide (nt) level, respectively, for the two pairs in the conserved regions), but RsMV28 had a 120-nt deletion in position 504 of the genome. Rhizoctonia solani mitovirus 22 encodes for an RdRp showing 80% identity with the ORF characterised from a dsRNA element isolated from a hypovirulent strain of *R. solani* in 1998 (Lakshman et al., 1998). Since this sequence is still annotated in the NCBI database, as a viral dsRNA element located in the fungal mitochondria, we decided to submit our sequence assigning it to a viral name. In previous work, the presence of ectopic DNA fragments derived from mitoviruses, infecting a fungal host (*Gigaspora margarita*), was detected but their function is still uncharacterised (Turina et al., 2018), contrary to analogous cDNA fragments found in insects and involved in anti-viral defence (Goic et al., 2016). Furthermore, RsMV22 is closely related to the viral dsRNA element isolated by Lakshman and co-authors in 1998, where a DNA stage was reported. To search for indications of mitovirus-derived DNA sequences in our samples, we analysed DNAs extracted from the eight *R. solani* isolates by qPCR using the same primers used for the detection of the viruses (by qRT-PCR) after the bioinformatic analysis**.** No evidence of amplification was observed (data not shown). We designed primers for PCR amplification of around 300 base pair fragments on the genome of RsMV22, RsMV21 and RsMV24. Also, in this case, we were able to amplify and clone the fragments in *E. coli* using the cDNA template and confirming the sequence, while no amplification was observed using extracted DNA as template. A 300 bp fragment was amplified and also cloned for RsOLV5, demonstrating the presence of the viral sequence only in the cDNA and not in the genomic DNA. Overall, we could not provide evidence of the existence of DNA fragments corresponding to the *R. solani* mitovirus in the isolates we tested.

The four viruses named Rhizoctonia solani ourmia-like virus 2 to 5 were grouped together with the fungal ourmia-like viruses, with Agaricus bisporus virus 15 as the closest hit in a BLAST search (Figure 2). Putative RdRp produced by RsOLV 2 and RsOLV 3 showed a 77% identity between them, and therefore, they are likely different isolates of the same species. A recent proposal grouped fungal ourmiaviruses together with plant ourmiaviruses in a new family called *Botourmiaviridae*; inside this family, three different genera containing fungal ourmiaviruses are established: *Botoulivirus*, *Magoulivirus* and *Scleroulivirus* (Figure 2). Our phylogenetic analysis shows that the new species identified in our study are part of a distinct clade containing also Agaricus bisporus virus 15, which appears to be basal to the proposed *Botourmiaviridae* family, therefore, posing the ground for a new virus family for which we propose the name Basidionarnaviridae, since currently it contains members infecting basidiomycetes.

2.4.2 Hepe-Virga group

Our bioinformatics pipeline unveiled four contigs encoding a single ORF showing similarity with viruses belonging to the *Endornaviridae* family. Phylogenetic analysis (Figure 3) on the predicted proteins shows that viral contigs are grouped together with viruses from the genus *Alphaendornavirus*, thus, we renamed the sequences as Rhizoctonia solani endornavirus 4 to 7. All the proteins predicted from the viral contigs show an RdRp domain, and proteins encoded from RsEV4, RsEV6, and RsEV7 also showed a helicase domain. A methyl-transferase domain was detected only in RsEV5 protein. In general, our endornavirus phylogenetic tree showed differences with the current taxonomic organisation of this family, that we think requires an update to recognise new genera inside the family. *Endornavirus* RsEV4, RsEV6 and RsEV7 constitute a new clade that could possibly turn into a new genus, for which we propose the name Gammaendornavirus (Figure 3).

A single 11.666-nt-long contig was identified as a new virus; an ORF prediction and BLAST analysis showed similarities with other characterised beny-like mycoviruses. Thus, the contig was named Rhizoctonia solani beny-like virus 1 (Figure 4). PCR with specific primers for RsBLV1 allowed us to amplify a 274 bp fragment from cDNA obtained from the infected isolates, while no specific amplification was observed on DNA extracted from the same isolate. Similarly to the still unpublished Sclerotium rolfsii beny-like virus 1, RsBLV1 encodes for one single ORF producing a 3584-amino acid-long protein showing a viral helicase domain and an RdRp domain. Other beny-like mycoviruses like Agaricus bisporus virus 8 and 13 show three ORFs that were not detected in RsBLV1.

Four contigs showed high similarity with viruses encoding ORFs with an RdRp domain belonging to the alphavirus supergroup (Wolf et al., 2018). Among these, a 2982-nt-long contig encodes for an ORF displaying an RdRp domain and a viral helicase domain. BLAST and phylogenetic analysis show that the closest viruses to this contig is Rhizoctonia solani flexivirus 2, and these two viruses probably belong to the newly characterised genus *Deltaflexivirus* (Figure 5). Viruses belonging to the genus *Deltaflexivirus* are supposed to encode small proteins in the 3-terminal part of the genome region and are usually around 8 kbp-long; all these characteristics were not observed in our contig or in the Rhizoctonia solani flexivirus 2 (Bartholomäus et al. 2017), but we cannot exclude that these two genomes are indeed partial. Taken together these considerations, we decided to name this virus Rhizoctonia solani flexi-like virus 1. The three remaining contigs belonged to viruses that we named Rhizoctonia solani alphavirus-like 1 to 3; RsAVL1 and RsAVL2 encode for uncomplete ORFs encoding RdRp, while the protein predicted from RsAVL3 seems to be a complete RdRp. Phylogenetic analysis (Figure 5) showed that these viruses are grouped together with three very small partial viral genomes discovered in *Rhizoctonia solani* (Rhizoctonia solani RNA virus 1 to 3) that have high homology with our sequences (Bartholomäus et al., 2017). Together with Sclerotinia sclerotiorum RNA virus L, these viruses form a distinct clade for which we propose a new family called Mycoalphaviridae. PCR amplification of overlapping fragments of around 800 bp from RsALV 3 confirmed the absence of the viral contig in the DNA of the host fungal isolate, confirming the viral nature of this segment.

2.4.3 Hypo related sequences

Six contigs showed a relationship to the *Hypoviridae* family (Figure 6). Among those, three contigs were closely related to the genus *Hypovirus*, while the remaining three were more similar to the still unclassified viruses generally called fusarivirus in a number of publications; identity percentages resulting from BLASTx analysis showed levels between 28.5% and 40.5% compared to the first hit. No other clearly correlated *R. solani* hypoviruses and fusariviruses were found in the literature. Thus, we named these contigs Rhizoctonia solani hypovirus 1 to 3 and Rhizoctonia solani fusarivirus 1 to 3. RsHV1 had an 18,371-bp-long genome. The ORF prediction showed just one large putative protein of 5344 amino acids where only one helicase domain can be detected. RdRp domains could not be observed and no GDD amino acid triplet, the hallmark of most viral RdRps, was found in the protein sequence. Nevertheless, from the BLAST analysis, it was clear that the RsHV1 protein had homology with the region encoding for the GDD domain in other hypoviruses like Sclerotinia sclerotiorum hypovirus 2 (for which an RdRp domain was annotated), even though such domain is not detected by common domain searching software, such as CDD sparkle and ExPASy-PROSITE (Sigrist et al., 2013, Marchler-Bauer et al., 2017). According to the International Committee on Taxonomy of Viruses - ICTV description, hypoviruses have a genome dimension of 9.1–12.7 kb. On the contrary, RsHV1 has one of the longest genomes known so far for a putative hypovirus. RsHV2 is a 9606-bp contig encoding for two ORFs, but the 3' proximal-ORF appears to be incomplete. Both ORFs did not show any conserved motif, although the protein amino acid sequence BLASTs with viruses belonging to the genus *Hypovirus* **.** RsHV3 is 5518-bp-long with an ORF prediction which underlined two putative proteins of which the 3'proximal ORF was likely an incomplete protein displaying a viral helicase domain. No RdRp domain could be detected in this case, and no conserved domains were observed on the 5' distal ORF. Alignments performed to produce a phylogenetic tree showed that RsHV1 and RsHV3 could be aligned on the helicase domain, and that the phylogenetic tree resulting from the analysis show a clade containing RsHV1, RsHV2, RsHV3 related to a hypovirus from *Agaricus bisporus* (Agaricus bisporus virus 2) and Sclerotinia sclerotiorum hypovirus 2. These viruses are included in a statistically well supported clade that separates them from the current characterised fusariviruses and the members of the genus *Hypovirus* (Figure 6). For this reason, we propose here the name of a new genus (Megahypovirus) for the large size of the genomes of the two complete sequences so far characterised in this group (SsHV2 and RsHV1). The presence of RsHV3 was confirmed through RT-PCR amplification; we were able to amplify and clone a fragment of the expected size just in the cDNA produced from the infected fungal isolate, and nothing was observed in the DNA extracted from the same fungus—confirming its viral nature and that the transcript was not derived from an endogenised viral fragment.

Among the three fusariviruses discovered (Figure 6), RsFV1 was the longest one, as it had a 10776-bp genome and the ORF prediction displayed four putative proteins. ORF 2 and 4 were the smallest: ORF 2 encoded a putative protein of 525 amino acids and ORF 4 encoded a putative protein of 578 amino acids. BLAST analysis and domain prediction could not find any convincing hit or conserved domain for both ORFs. ORF 1 was 731

amino acids long and it had a viral helicase domain, while the RdRp domain was found in ORF 3, together with another helicase domain. The same genome organisation was observed for RsFV2, a 10710-bp contig showing a PolyA site at the 3' of the sequence and encoding for four putative proteins. In this virus, ORF 2 and 4 (472 and 718 amino acids, respectively) were short ORFs with no significant homology with other viral proteins. ORF1 had a viral helicase domain and ORF 3 had a helicase and an RdRp domain. Finally, RsFV3 was a 5959-bp contig encoding for just one protein. The predicted protein sequence showed helicase and RdRp domains as observed for the other two fusarivirus and a PolyA site at the 3' end. As previously described, fusariviruses often encoded two ORFs, with the 5'-ORF encoding for the RdRp. The new viruses that we discovered in this work presented some different characteristics that are peculiar for this group of viruses (four ORFs in RsFV1 and RsFV2). We decided to further investigate the ORF prediction of RsFV1 and RsFV2 by designing PCR primers and amplifying sequences overlapping the regions of discontinuity between the four ORFs to verify possible mistakes in the sequence. Cloned fragments were sequenced and compared to the viral genomes proving that the reference sequence assembled from Trinity was identical to the one amplified through PCR. Thus, the four proteins predicted cannot be due to an error in the RNAseq assembly. Phylogenetic analysis confirmed BLAST analysis and placed our three fusariviruses together with others already characterised in the same clade. We propose that currently recognised fusariviruses, based on their sequence length and genome organisation, be subdivided into at least two further genera (Figure 6).

2.4.4 Bunya-Arena like sequences

We identified a contig encoding for a protein showing the RdRp domain from bunyaviruses and we named this contig Rhizoctonia solani bunya/phlebo-like virus 1. RsBPLV1 is 7804 bp long and the putative RdRp is 2513 amino acids long. BLASTp analysis on the RdRp showed 30.59% identity with the closest virus in the database. Phylogenetic analysis previously carried out showed that the fungal negative single-strand RNA viruses belong to three orders: the order *Mononegavirales* and family *Mymonaviridae* (Liu et al., 2014), the order *Serpentovirales* which ~~the second~~ likely includes two families that we propose to name Alphamycoserpentoviridae and Betamycoserpentoviridae, and the order *Bunyavirales* that includes other mycoviruses. RsBPLV1 was clearly in this order, in our analysis (Figure 7). Nevertheless, two distinct new families that include mycoviruses can be proposed inside this order: for the one that includes RsBPLV1 we propose the name Mycophleboviridae, whereas for the other well supported clade, we propose the name Mycobunyaviridae.

2.4.5 Partiti-Picobirna sequence group

Contigs coding for proteins showing high homology with partitivirus RdRps were detected in our samples (Figure 8). From the initial BLAST analysis, Four contigs encoding for partitivirus RdRp and three contigs encoding for partitivirus coat proteins were retrieved. Among these seven contigs, one encoding for an RdRp and one encoding for a coat protein were easily matched as part of the same virus since they showed almost complete identity with an already characterised virus belonging to the genus *Alphapartitivirus* called Rhizoctonia solani dsRNA virus 2. Indeed, these contigs were always detected in

the same fungal isolates through qRT-PCR, confirming that these two segments belonged to the same virus. The remaining contigs were submitted to the NCBI database as Rhizoctonia solani partitivirus 6 to 8. RsPV7 and RsPV8 RdRps were part of the genus *Alphapartitiviruses*, while RsPV6 belonged to the group hosting viruses from the genus *Betapartitivirus* (Figure 8). A correct correlation between RdRp and coat protein of RsPV6 and RsPV7 have been complicated by the fact that the two viruses were found in the same fungal isolate in our collection. Thus, we grouped together the two RNAs attributed to RsPV7 because the RdRp protein and the coat protein produced from the two segments had as a first BLAST hit the same two proteins from a single viral species called Trichoderma atroviride partitivirus 1 (Table 1**)**. No RNA2 producing coat protein was detected initially for RsPV8, but we tried to look for the missing genome segment by a direct TBLASTn search on the RNAseq assembly using as a query the coat protein sequence from the virus whose RdRp is more similar to RsPV8 RdRp. We were able to find a short contig producing a protein that showed homology with partitivirus coat proteins, however, the coverage of this contig is low and attempts to extend the sequence with MITObim were not successful. Nevertheless, qRT-PCR analysis confirmed the presence of the contig only in the fungal isolate hosting the RsPV8 RNA1 fragment. Taken together, we submitted to Genbank the contig as RsPV8's partial RNA2.

Two short contigs of 1827 and 1888 bp were found as part of the same viral genome, where the 1827-bp contig encoded a single incomplete ORF producing a protein showing homology with RdRp from viruses with a bipartite genome related to a partitivirus that we have previously described (Nerva et al., 2016), and provisionally named bipartite viruses. The 1888-bp segment codes for a hypothetical protein, showing homology with the same virus group. Phylogenetic analysis (Figure 8) placed this virus basal to the virus group already proposed in literature as bipartite (Nerva et al., 2016), thus, we decided to rename this virus Rhizoctonia solani bipartite-like virus 1.

2.4.6 Toti-Chryso

Five contigs encoded for putative RdRp with similarities to viruses belonging to the Toti-Chryso group (Figure 8). These contigs were renamed Rhizoctonia solani dsRNA virus 6 to 10, and phylogenetic analysis grouped these viruses together in a clade containing unclassified viruses related to the genus *Megabirnavirus*. This clade has already been proposed to form a new genus called Phlegivirus by previous work (Petrzik et al., 2016)**.** Here, we show that such taxon comprises two genera, and therefore, we propose the family Phlegiviridae to include both of them (Figure 8). Rhizoctonia solani dsRNA 6 and 10 showed two ORFs; the first was a protein of unknown function, and the second ORF encoded for the RdRp. The RsdsRNA6 genome was 11847 bp long: the first ORF was 2300 amino acids long and the RdRp was a smaller protein of 1125 amino acids. RsdsRNA 10 was 9416 nt-long. The first ORF was 1636 amino acids long, while the predicted RdRp was a 1475-amino acid protein. No clear conclusion about genome organisation of RsdsRNA 7 to 9 could be drawn since all three genomic segments encoded for a single ORF, the putative RdRps. Furthermore, RsdsRNA 9 ORF was partial, while for the other two ORFs, a stop codon could be detected. Since these three contigs were much shorter compared to the closest viruses according to the phylogenetic analysis, we can assume that they are

probably incomplete. So far, attempts to extend the viral contigs were unsuccessful. PCR amplification of a target fragment from RsdsRNA10 gave positive results only when using the cDNA of the infected isolate as a template, while no PCR amplification was observed from total DNA, once again suggesting that the RNA was not derived from transcription of endogenised viral fragments.

2.4.7 ORFans fragments

Four fragments resulting from our bioinformatic pipeline were detected as viral but could not be located in any known taxonomical group, and no evidence of an RdRp domain was detected in these sequences. We called these contigs Rhizoctonia solani putative virus 1 to 4. RsPuV 1 was a 6311-bp-long sequence, and the ORF encoded showed a conserved viral helicase domain, which has similarities with viruses from different groups like *Tymovirus* and *Endornaviruses*. Rhizoctonia solani putative virus 2 was 7137 bp long, and in this case, the 2083-amino acid long protein only had a conserved viral helicase motif. Rhizoctonia solani putative virus 3 had a 7713-bp sequence, encoding for a 2011-amino acid long uncomplete protein, showing a viral methyl transferase domain. BLAST analysis showed little homology with viruses from the genus *Tymovirus*. Rhizoctonia solani putative virus 4 had a 7833-bp sequence coding for an ORF resulting in a putative 2414-amino acid protein with a viral helicase domain. In each of these sequences, the homology with other viruses is always located on the small part of the protein showing the conserved domain (helicase and methyltransferase); thus, it is hard to hypothesise a specific taxonomic placement for these viral sequences. Indications that these fragments are of viral origin also comes from the fact that no DNA was detected corresponding to these fragments using DNA as template for PCR amplification (Supplementary Figure 2, online). We confirmed the presence of these putative virus RNA fragments through RT-PCR using cDNA as a template, which resulted in bands of the expected size. We cannot exclude that some of these viral fragments are part of a multipartite virus that has escaped our detection or are associated with one of the viruses we have described in this paper.

2.5 Conclusions

In the present study we reported sequences corresponding to mostly new viral species belonging to the positive sense ssRNA genome virus groups (*Mitovirus*, *Botourmiaviridae*, *Hypovirus*, *Endornavirus*, Hepe-Virga-like), and to dsRNA virus groups such as the *Alphapartitivirus* and *Gammapartitivirus*. Furthermore, a single negative strand virus in the *Bunyavirales* was also characterised in our collection.

The number of viral species infecting each *R. solani* isolates ranged from 6 (IBRS 15) to 27 (IBRS23). The mitoviruses RsMV24 and RsMV29, and the endornavirus RsEV7, were the most prevalent viruses since they occurred in all the isolates. However, some mycoviruses were detected in only one isolate of *R. solani*: RsMV21, RsMV23, RsMV25, RsMV26, RsMV27, RsMV28, RsMV30, RsMV31, RsMV36, RsEV6, RsPV6, RsPV7, RsPV8, RsBPLV1, RsBLV1, RsBeLV1, RsALV3, RsOLV5, RsPuV4, RsFV1, RsHV2 and RsHV3.

Successful uses of mycoviruses have already been reported for the biological control of *Cryphonectria parasitica* (Nuss, 1992) in natural conditions. Some promising hypovirulent strains were also detected for *Ophiostoma novo-ulmi* (Hong et al, 1999), *Fusarium graminearum* (Chu et al., 2002), *Botrytis cinerea* (Wu et al., 2007) and *Rosellinea necatrix* (Chiba et al., 2009).

The ability of fungi to cause disease in the host plant seems to not be affected by the wide diversity of the viral species detected in the same *Rhizoctonia* isolates, since all the isolates induced the same kind of symptoms in the *Z. japonica* host. Indeed, in greenhouse tests, the isolates IBRS11, IBRS19 and IBRS23 were the most pathogenic, while the isolate IBRS15 was the least pathogenic (Picarelli et al., in press). We could not correlate virus distribution to the geographical distribution of the isolates (Supplementary Figure 1, online), and it is surprising that the comprehensive characterisation of the three viromes associated with *R. solani* so far have only a minimal overlap of identical sequences (Bartholomäus et al., 2017, Marzano et al., 2016). Nevertheless, some new taxonomical groups that include mostly *R. solani* viruses were common to the various studies as was the case of some subclades in the *Mitovirus* genus, the proposed Gammaendornavirus genus, the proposed Mycoalphaviridae family and the proposed Phlegiviridae family. Our study helps to define two completely new clades of mycoviruses, the proposed genus Megahypovirus and the family Basidiourmiaviridae.

2.6 Author contributions

MP collected the samples, characterized the fungal isolates, checked their virulence, and extracted dsRNA. MF extracted the total RNA, carried out all the virus annotation, carried out RT-PCR and PCR, mapped reads on the genome, carried out the phylogenetic analysis, and wrote the manuscript. ER curated virus nomenclature and classification, and carried out some phylogenetic analyses and edited the manuscript. LN carried out the initial bionformatic virome characterization. MC extended the genome segments with MitoBIM. MT edited the manuscript and supervised LN, MC, and MF. AC supervised MP, planned the experiments and edited the manuscript and deposited sequences in the databases with ER.

2.7 Figures and Tables

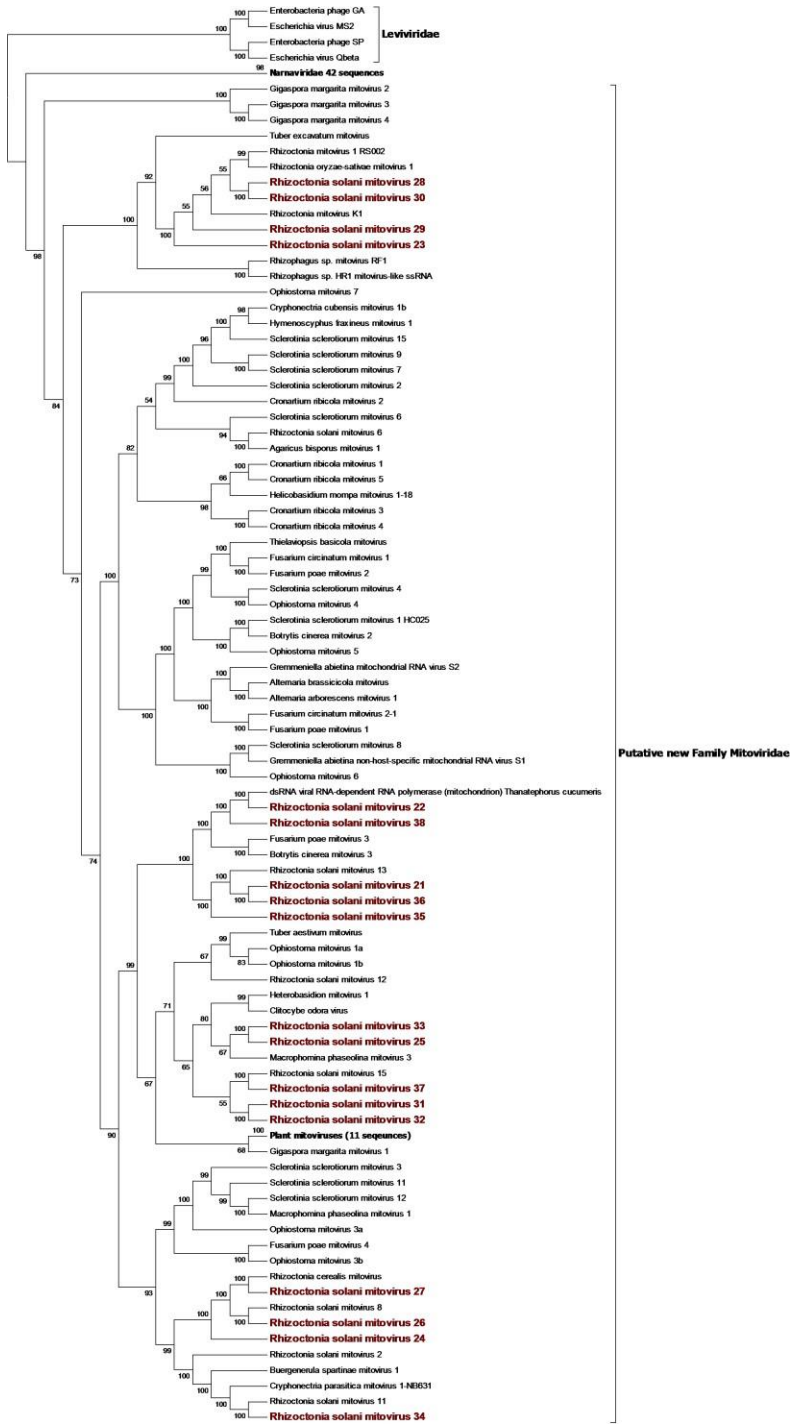


Figure 1. Phylogenetic analysis of positive sense RNA viruses related to the genus *Mitovirus*. 142 sequences have been used to produce an alignment starting from viruses belonging to the family *Narnaviridae*, *Botourmiaviridae,* and *Leviviridae* as outgroup; the phylogenetic tree was built using the maximum likelihood method, the best choice for the substitution model according to ModelFinder was PMB+F+I+G4. Ultrafast bootstrap analysis was performed with 1000 replicates, and branches displaying values below 50 were collapsed. Viruses discovered in this work are in bold red ink; 42 sequences belonging to the *Narnaviridae* family have been compressed in one branch. A list of accession numbers of the sequences used for this analysis can be found in Supplementary on line table 3. Detailed information about softwares used for the analysis can be found in the material and methods section.

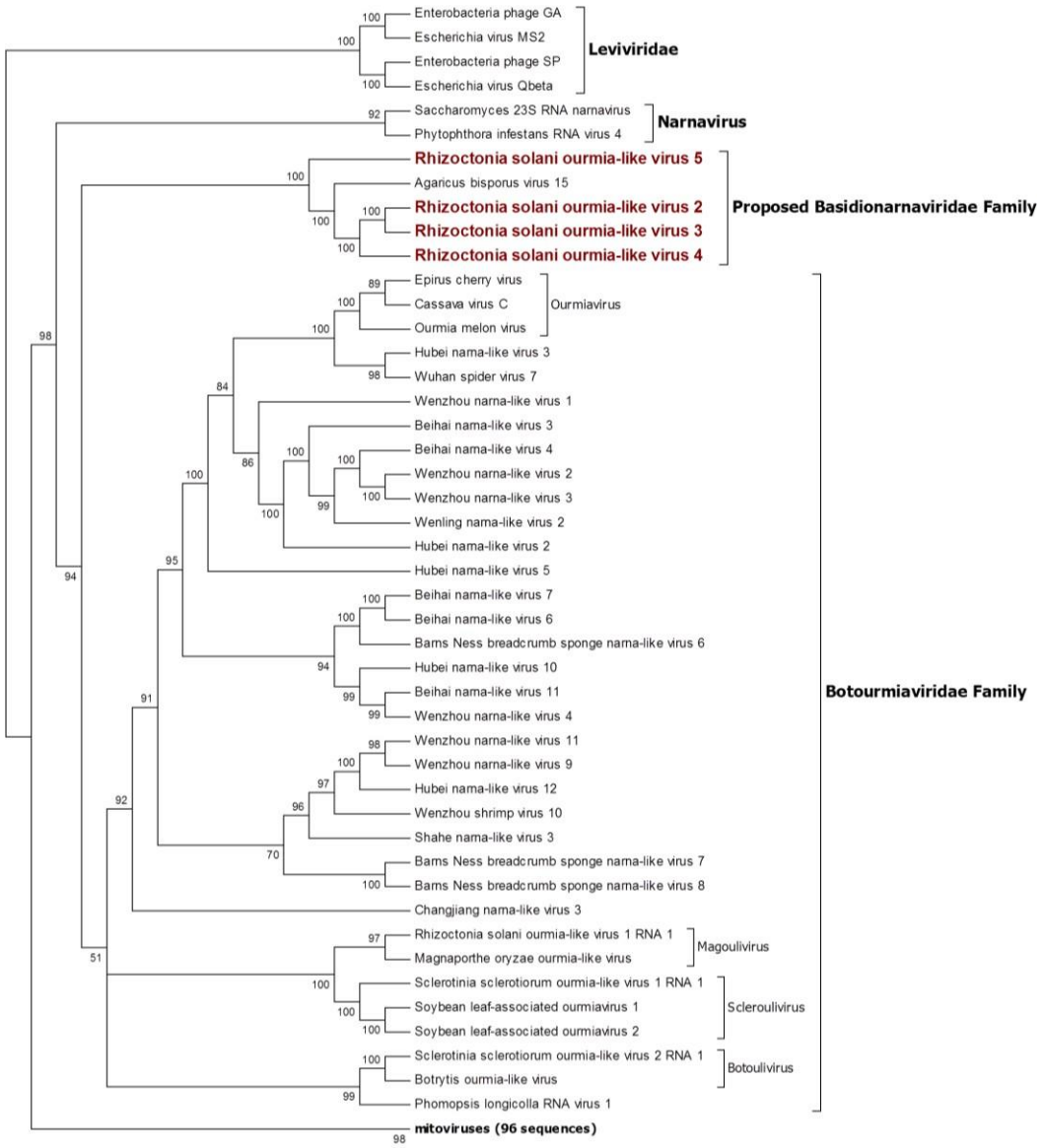

Enterobacteria phage GA
Escherichia virus MS2
Enterobacteria phage SP
Escherichia virus Qbeta
Leviviridae
Saccharomyces 23S RNA narnavirus
Phytophthora infestans RNA virus 4
Narnavirus
Rhizoctonia solani ourmia-like virus 5
Agaricus bisporus virus 15
Rhizoctonia solani ourmia-like virus 2
Rhizoctonia solani ourmia-like virus 3
Rhizoctonia solani ourmia-like virus 4
Proposed Basidionarnaviridae Family
Epirus cherry virus
Cassava virus C
Ourmia melon virus
Ourmiavirus
Hubei narna-like virus 3
Wuhan spider virus 7
Wenzhou narna-like virus 1
Beihai narna-like virus 3
Beihai narna-like virus 4
Wenzhou narna-like virus 2
Wenzhou narna-like virus 3
Wenling narna-like virus 2
Hubei narna-like virus 2
Hubei narna-like virus 5
Beihai narna-like virus 7
Beihai narna-like virus 6
Barns Ness breadcrumb sponge narna-like virus 6
Hubei narna-like virus 10
Beihai narna-like virus 11
Wenzhou narna-like virus 4
Wenzhou narna-like virus 11
Wenzhou narna-like virus 9
Hubei narna-like virus 12
Wenzhou shrimp virus 10
Shahe narna-like virus 3
Barns Ness breadcrumb sponge narna-like virus 7
Barns Ness breadcrumb sponge narna-like virus 8
Changjiang narna-like virus 3
Botourmiaviridae Family
Rhizoctonia solani ourmia-like virus 1 RNA 1
Magnaporthe oryzae ourmia-like virus
Magoulivirus
Sclerotinia sclerotiorum ourmia-like virus 1 RNA 1
Soybean leaf-associated ourmiavirus 1
Soybean leaf-associated ourmiavirus 2
Scleroulivirus
Sclerotinia sclerotiorum ourmia-like virus 2 RNA 1
Botrytis ourmia-like virus
Botoulivirus
Phomopsis longicolla RNA virus 1
mitoviruses (96 sequences)

Figure 2. Phylogenetic analysis of positive sense RNA viruses related to the genus *Botourmiaviridae*. 142 sequences have been used to produce an alignment starting from viruses belonging to the family *Narnaviridae*, *Botourmiaviridae,* and *Leviviridae* as outgroup; phylogenetic tree was built using the maximum likelihood method the best choice for the substitution model according to ModelFinder was PMB+F+I+G4. Ultrafast bootstrap analysis was performed with 1000 replicates, and branches displaying values below 50 were collapsed. Viruses discovered in this work are in bold red ink; 96 sequences belonging to the genus *Mitovirus* have been compressed in one branch. A list of accession numbers of the sequences used for this analysis can be found in Supplementary on line table 3. Detailed information about softwares used for the analysis can be found in the material and methods section.

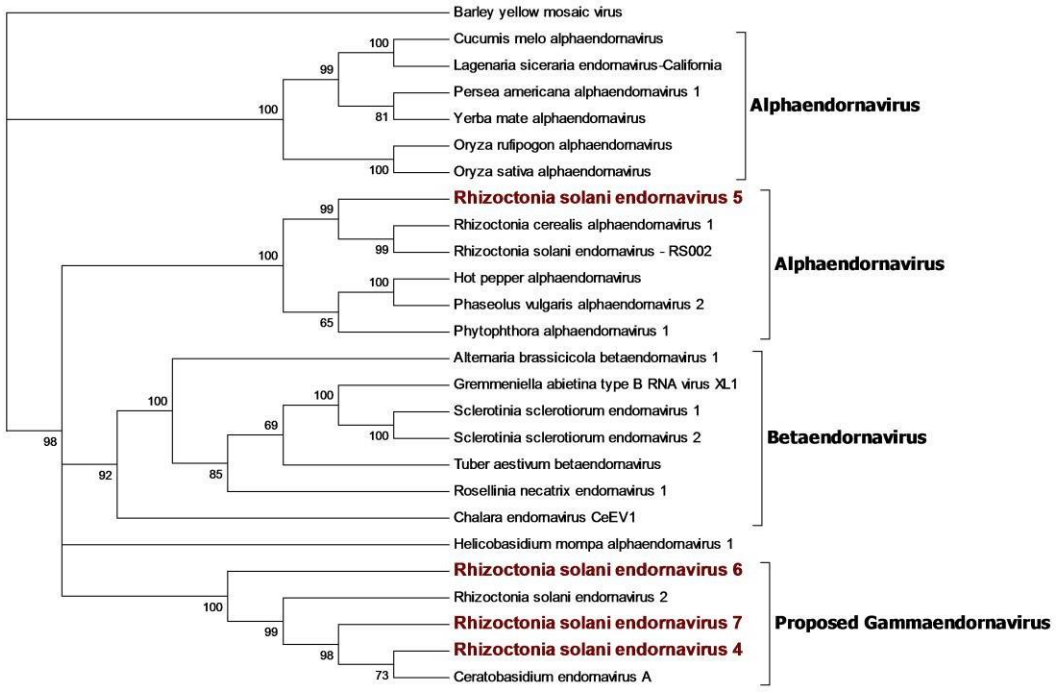


Figure 3. Phylogenetic analysis of viruses belonging to the family *Endornaviridae*. Twenty-six viral sequences were used to build the alignment. A phylogenetic tree was built using the maximum likelihood method and the best substitution model selected by ModelFinder was VT+F+I+G4. Ultrafast bootstrap analysis was performed with 1000 replicates and branches displaying values below 50 were collapsed. Viruses discovered in this work are outlined in bold red. Barley yellow mosaic virus, belonging to the family *Potyviridae*, was used as outgroup. A list of accession numbers of the sequences used for this analysis can be found in Supplementary on line table 4. Detailed information about softwares used for the analysis can be found in the material and methods section.

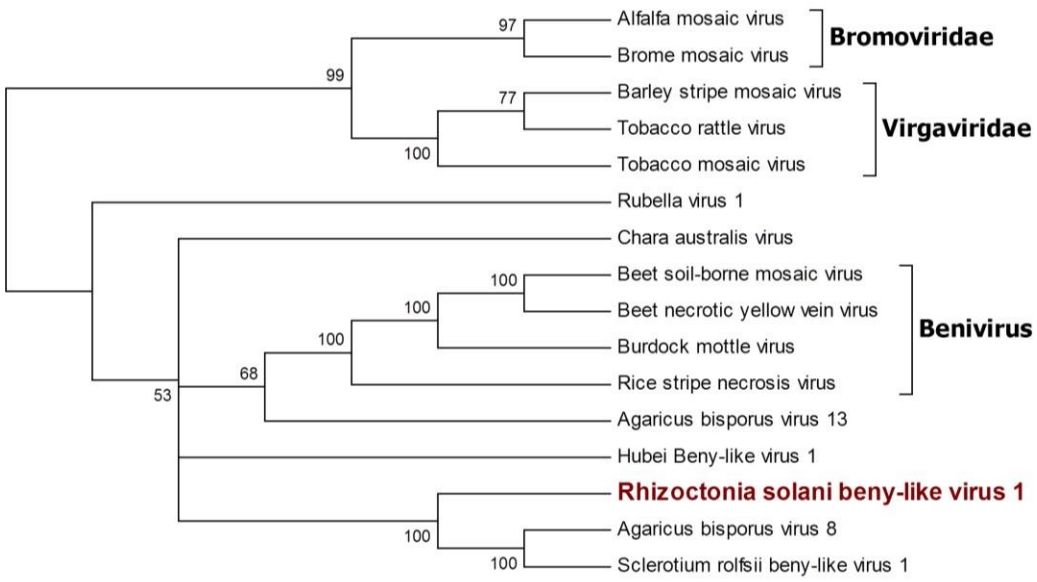


Figure 4. Phylogenetic analysis of viruses belonging to the family *Benyviridae*. Sixteen viral sequences were aligned, and a phylogenetic tree was derived using the maximum likelihood method. The best substitution model selected by ModelFinder was VT+F+I+G4. Ultrafast bootstrap analysis was performed with 1000 replicates, and branches displaying values below 50 were collapsed. Viruses discovered in this work are outlined in bold red. *Bromoviridae* and *Virgaviridae* were used as outgroups. A list of accession numbers of the sequences used for this analysis can be found in Supplementary on line table 4. Detailed information about softwares used for the analysis can be found in the material and methods section.

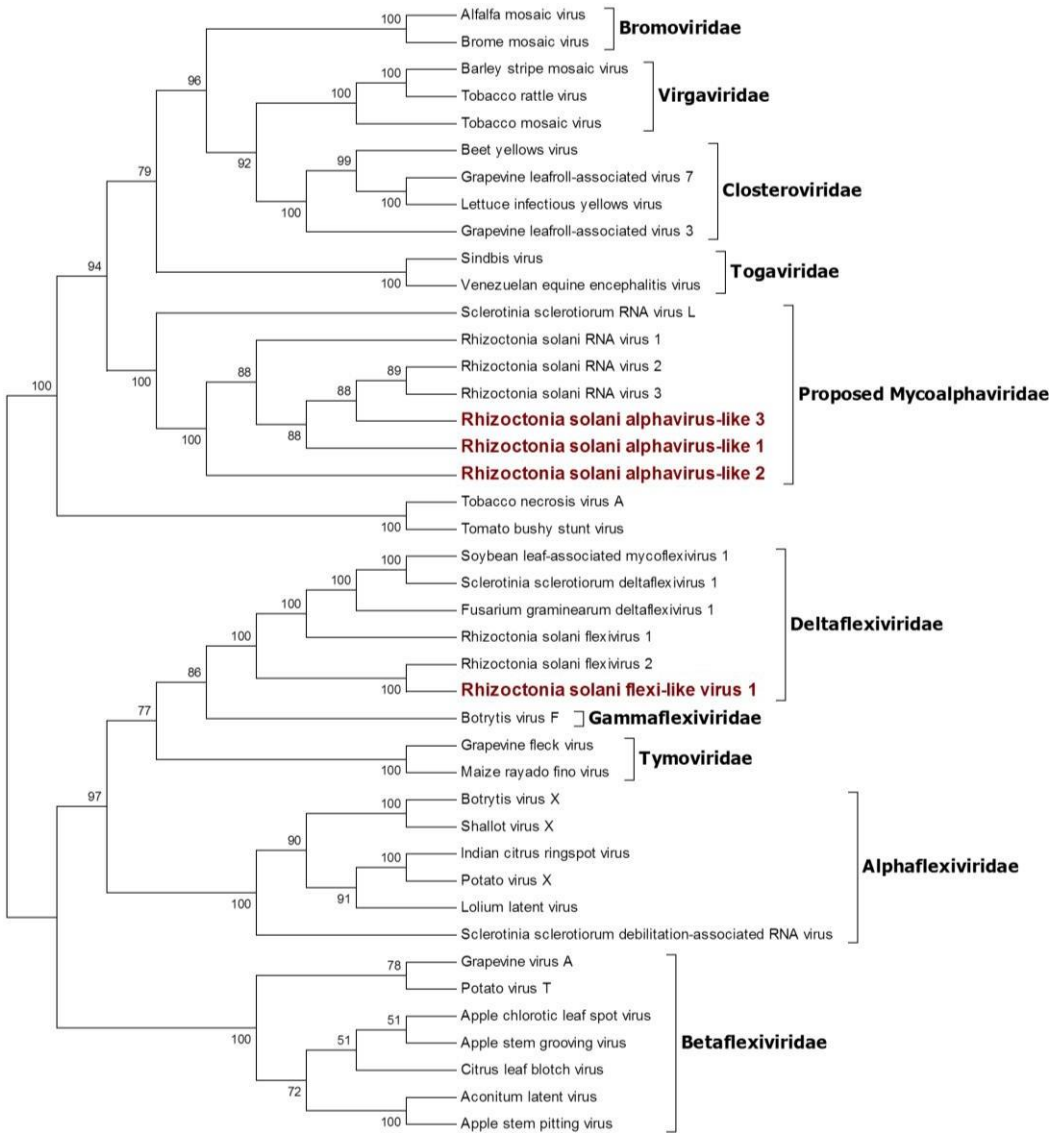


Figure 5. Phylogenetic analysis of viruses related to the *Alphavirus* supergroup. Forty-two sequences were aligned, and a phylogenetic tree was built using the maximum likelihood method. The best substitution model selected by ModelFinder was VT+F+I+G4 and ultrafast bootstrap analysis was performed with 1000 replicates. Branches displaying values below 50 were collapsed. Viruses discovered in this work are outlined in bold red ink. A list of accession numbers of the sequences used for this analysis can be found in Supplementary on line table 4. Detailed information about softwares used for the analysis can be found in the material and methods section.

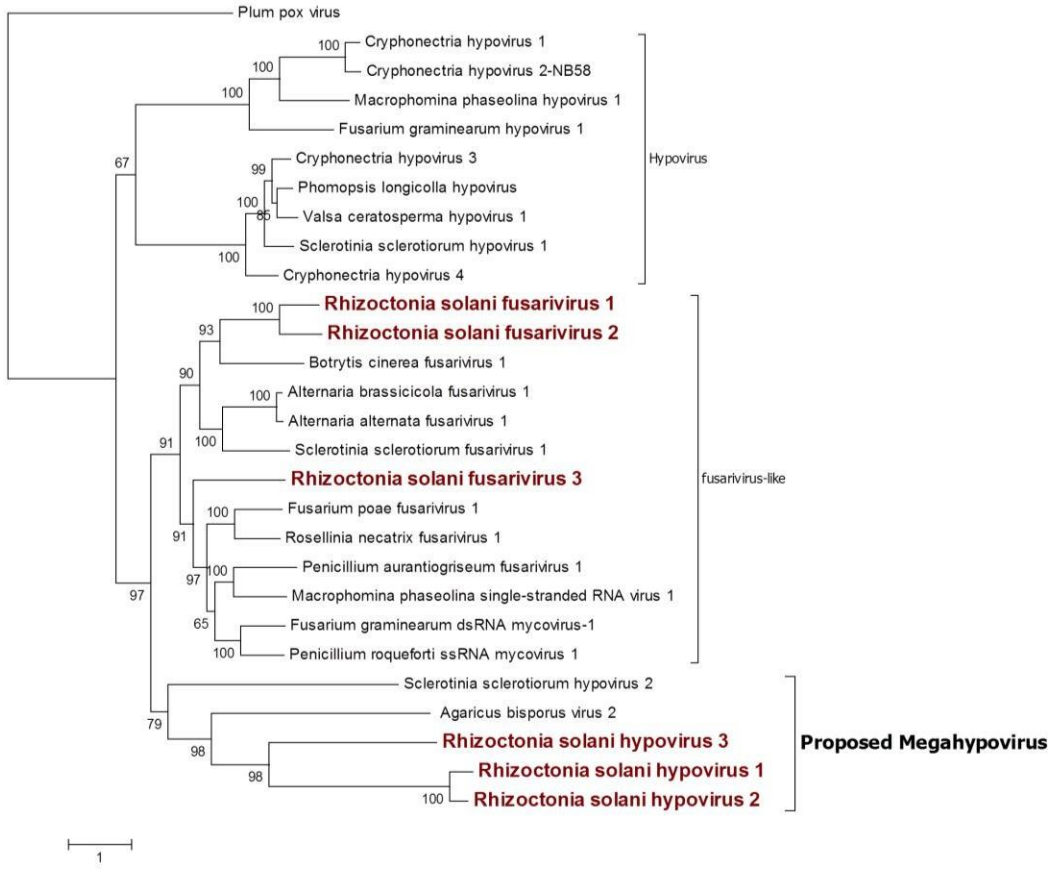


Figure 6. Phylogenetic analysis of viruses related to the family *Hypoviridae*. Twenty-eight sequences were used to build an alignment and to derive a phylogenetic tree using the maximum likelihood method. The best substitution model selected by ModelFinder was VT+F+G4 and an ultrafast bootstrap analysis was performed with 1000 replicates. Branches displaying values below 50 were collapsed. Viruses discovered in this work are outlined in bold red. Plum pox virus, belonging to the family *Potyviridae*, was used as outgroup. A list of accession numbers of the sequences used for this analysis can be found in Supplementary on line table 5. Detailed information about softwares used for the analysis can be found in the material and methods section.

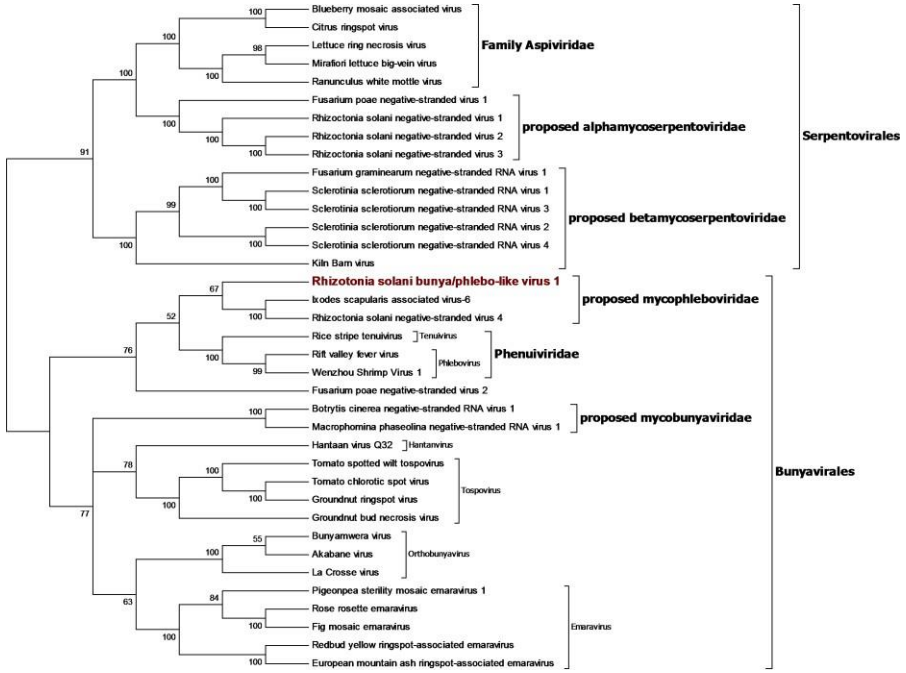


Figure 7. Phylogenetic analysis of viruses related to the orders *Bunyavirales* and *Serpentovirales*. Thirty-seven sequences were aligned, and the best substitution model selected by ModelFinder was VT+F+G4 and implemented in a maximum likelihood method to derive the Phylogenetic tree. Ultrafast bootstrap analysis was performed with 1000 replicates and branches displaying values below 50 were collapsed. The Virus discovered in this work is outlined in red. A list of accession numbers of the sequences used for this analysis can be found in Supplementary on line table 6**.** Detailed information about software used for the analysis can be found in the material and methods section.

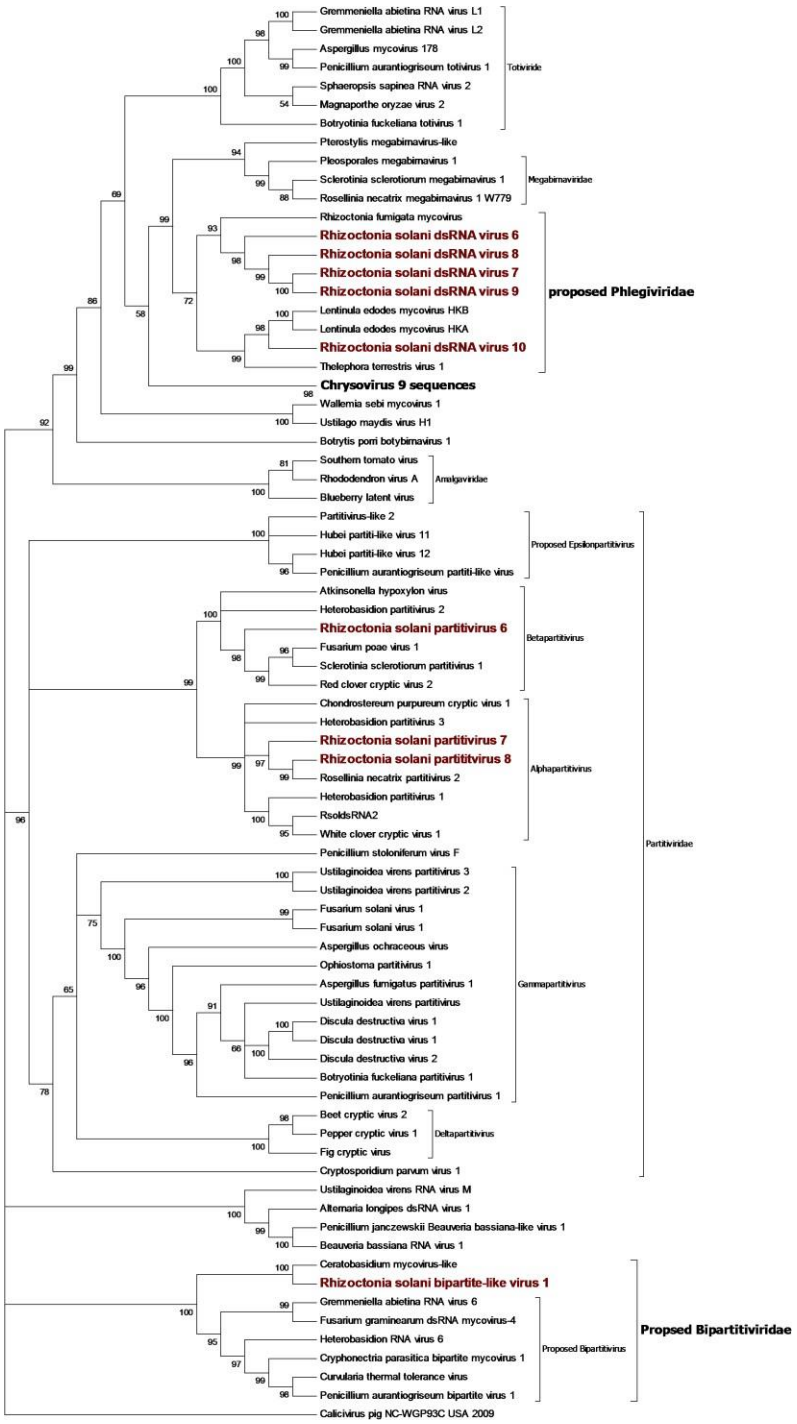


Figure 8. Phylogenetic analysis of dsRNA viruses. Eighty-four sequences were used to build an alignment. Phylogenetic tree was built using the maximum likelihood method,

the best substitution model selected by ModelFinder was Blosum62+F+G4 and ultrafast bootstrap analysis was performed with 1000 replicates, branches displaying values below 50 were collapsed. Viruses discovered in this work are underlined in red. A list of accession numbers of the sequences used for this analysis can be found in Supplementary table 7. Detailed information about software used for the analysis can be found in the material and methods section.

Table 1. List of viruses discovered in *Rhizoctonia solani* isolates from Sao Paulo State (Brazil). For each virus the NCBI code is reported together with the abbreviation used in the paper, segment length, number of reads mapping the segment and first hit obtained from BLASTx analysis with query cover and identity percentages.

Virus Name		NCBI ID	Abbreviation	Segment length (bp)	N. of reads	BLASTx first hit	Ident	Query cover
Rhizoctonia solani endornavirus 4		MK393902	RsEV4	20215	23587	Endornavirus-like virus	35.16	23
Rhizoctonia solani endornavirus 5		MK393903	RsEV5	16227	6414	Rhizoctonia cerealis alphaendornavirus 1	47.31	36
Rhizoctonia solani endornavirus 6		MK393904	RsEV6	15273	4441	Morchella importuna endornavirus 2	45.18	49
Rhizoctonia solani endornavirus 7		MK393905	RsEV7	14325	73559	Rhizoctonia solani endornavirus 2	30.69	83
Rhizoctonia solani partitivirus 8	RNA1	MK532273	RsPV8	1958	2239	Rhizoctonia oryzae-sativae partitivirus 1	53.19	90
	RNA2	MK532274		778	178	Heterobasidion partitivirus 20	42.35	74
Rhizoctonia solani partitivirus 6	RNA1	MK507781	RsPV6	2401	5189	Fusarium poae partitivirus 2	50.96	90
	RNA2	MK507782		2310	11867	Rosellinia necatrix partitivirus 8	58.08	84
Rhizoctonia solani partitivirus 7	RNA1	MK507783	RsPV7	1912	10488	Trichoderma atroviride partitivirus 1	68.08	95
	RNA2	MK507784		1867	3218	Trichoderma atroviride partitivirus 1	31.69	86
Rhizoctonia solani dsRNA virus 2	RNA1	MK400668	RsdsRNA2	1942	32967	Rhizoctonia solani dsRNA virus 2	99.36	96
	RNA2	MK400669		1727	16114	Rhizoctonia solani dsRNA virus 2	98.77	84
Rhizoctonia solani bipartite-like virus 1	RNA1	MK492913	RsBLV1	1827	279	Ceratobasidium mycovirus-like	71.50	69
	RNA2	MK492914		1888	331	Ceratobasidium mycovirus-like	61.60	41
Rhizoctonia solani dsRNA virus 6		MK507788	RsdsRNA6	11847	5257	Rhizoctonia fumigata mycovirus	30.95	20
Rhizoctonia solani dsRNA virus 7		MK507789	RsdsRNA7	3523	1839	Rhizoctonia fumigata mycovirus	31.70	78
Rhizoctonia solani dsRNA virus 8		MK507790	RsdsRNA8	2905	1370	Rhizoctonia fumigata mycovirus	30.83	85

Rhizoctonia solani dsRNA virus 9	MK507791	RsdsRNA9	2162	445	Rhizoctonia fumigata mycovirus	35.48	70
Rhizoctonia solani dsRNA virus 10	MK532272	RsdsRNA10	9416	3945	Sclerotium rolfsii mycovirus dsRNA 1	37.37	37
Rhizoctonia solani beny-like virus 1	MK507778	RsBeLV1	11666	11135	Sclerotium rolfsii beny-like virus 1	36.35	39
Rhizoctonia solani bunya/phlebo-like virus 1	MK507779	RsBPLV1	7804	25865	Barns Ness serrated wrack bunya/phlebo-like virus 1	30.56	60
Rhizoctonia solani flexi-like virus 1	MK507787	RsFLV1	2982	1848	Rhizoctonia solani flexivirus 2	70.50	27
Rhizoctonia solani alphavirus-like 1	MK507793	RsALV1	2414	949	Rhizoctonia solani RNA virus 3	82.14	38
Rhizoctonia solani alphavirus-like 2	MK507792	RsALV2	3396	509	Rhizoctonia solani RNA virus 1	69.38	27
Rhizoctonia solani alphavirus-like 3	MK507786	RsALV3	6752	924	Rhizoctonia solani RNA virus 3	83.12	13
Rhizoctonia solani mitovirus 21	MK372892	RsMV21	4100	2008721	Rhizoctonia solani mitovirus 13	51.13	66
Rhizoctonia solani mitovirus 22	MK490928	RsMV22	2177	2709274	dsRNA viral RdRp (mitochondrion) [Thanatephorus cucumeris]	80.15	73
Rhizoctonia solani mitovirus 23	MK375261	RsMV23	2792	16946	Ceratobasidium mitovirus A	61.98	83
Rhizoctonia solani mitovirus 24	MK372893	RsMV24	2149	752995	Macrophomina phaseolina mitovirus 2	85.48	75
Rhizoctonia solani mitovirus 25	MK372894	RsMV25	3767	24180	Macrophomina phaseolina mitovirus 3	38.49	43
Rhizoctonia solani mitovirus 26	MK372895	RsMV26	2580	94168	Rhizoctonia cerealis mitovirus	46.15	90
Rhizoctonia solani mitovirus 27	MK372896	RsMV27	3176	11379	Rhizoctonia cerealis mitovirus	50.42	77
Rhizoctonia solani mitovirus 28	MK372897	RsMV28	2640	2693	Rhizoctonia mitovirus 1	39.85	76
Rhizoctonia solani mitovirus 29	MK372898	RsMV29	2904	56286	Binucleate Rhizoctonia mitovirus K1	42.14	79
Rhizoctonia solani mitovirus 30	MK372899	RsMV30	2760	3461	Binucleate Rhizoctonia mitovirus K1	42.74	80
Rhizoctonia solani mitovirus 31	MK372900	RsMV31	3820	1605182	Rhizoctonia solani mitovirus 7	52.80	69
Rhizoctonia solani mitovirus 32	MK372901	RsMV32	3409	1432655	Rhizoctonia solani mitovirus 7	42.81	73

Rhizoctonia solani mitovirus 33	MK372902	RsMV33	2733	505184	Rhizoctonia solani mitovirus 7	37.97	69
Rhizoctonia solani mitovirus 34	MK372903	RsMV34	3389	10356	Rhizoctonia solani mitovirus 11	44.00	71
Rhizoctonia solani mitovirus 35	MK490929	RsMV35	3772	2459926	Rhizoctonia solani mitovirus 13	43.70	67
Rhizoctonia solani mitovirus 36	MK490930	RsMV36	2562	958130	Rhizoctonia solani mitovirus 13	54.90	71
Rhizoctonia solani mitovirus 37	MK372904	RsMV37	3597	379328	Rhizoctonia solani mitovirus 15	51.06	69
Rhizoctonia solani mitovirus 38	MK372905	RsMV38	3197	2467420	dsRNA viral RdRp (mitochondrion) [Thanatephorus cucumeris]	45.42	66
Rhizoctonia solani ourmia-like virus 2	MK372906	RsOLV2	4104	22042	Rhizoctonia solani ourmia-like virus 1 RNA 1	78.18	44
Rhizoctonia solani ourmia-like virus 3	MK372907	RsOLV3	3223	3795	Rhizoctonia solani ourmia-like virus 1 RNA 1	76.59	57
Rhizoctonia solani ourmia-like virus 4	MK372909	RsOLV4	4557	5471	Rhizoctonia solani ourmia-like virus 1 RNA 1	48.96	34
Rhizoctonia solani ourmia-like virus 5	MK372908	RsOLV5	5234	3868	Agaricus bisporus virus 15	25.89	29
Rhizoctonia solani fusarivirus 1	MK558257	RsFV1	10776	2788	Rosellinia necatrix fusarivirus 2	40.00	25
Rhizoctonia solani fusarivirus 2	MK558256	RsFV2	10710	36543	Rosellinia necatrix fusarivirus 2	40.57	25
Rhizoctonia solani fusarivirus 3	MK558258	RsFV3	5959	8858	Fusarium graminearum dsRNA mycovirus-1	38.09	57
Rhizoctonia solani hypovirus 1	MK558259	RsHV1	18371	86930	Sclerotium rolfsii hypovirus 1	27.35	32
Rhizoctonia solani hypovirus 2	MK558260	RsHV2	9606	4582	Sclerotium rolfsii hypovirus 1	27.16	24
Rhizoctonia solani hypovirus 3	MK558255	RsHV3	5518	1553	Agaricus bisporus virus 2	28.49	18
Rhizoctonia solani putative virus 1	MK507780	RsPuV1	6311	56781	Lily symptomless virus	24.91	9

Rhizoctonia solani putative virus 2	MK507785	RsPuV2	7137	18568	Sanxia atyid shrimp virus 1	22.73	9
Rhizoctonia solani putative virus 3	MK532275	RsPuV3	7214	3662	Guarapuava tymovirus-like 1	31.91	9
Rhizoctonia solani putative virus 4	MK507793	RsPuV4	7833	7495	Gayfeather mild mottle virus	25.28	9

Table2. List of Ct (threshold cycles) obtained from qRT-PCR analysis for each viral genome segment in every fungal isolate investigated. Empty boxes have been left where no specific amplification curve was detected.

Viral contig		***Rhizoctonia solani* isolates**							
		IBRS22	**IBRS23**	**IBRS04**	**IBRS07**	**IBRS11**	**IBRS15**	**IBRS16**	**IBRS19**
RsEV4					27			26	28
RsEV5		28	31	25	28			30	
RsEV6									27.8
RsEV7		25	24	22	24	24	24	23	24
RsPV6	RNA1		22						
	RNA2		19.4						
RsPV7	RNA1		24						
			27						
RsPV8	RNA1							23	
	RNA2							27	
RsdsRNA2	RNA1	24	29	22	27			29	24
	RNA2	29	30	21	26			25	23
RsBLV1	RNA1					28			
	RNA2					28.5			
RsdsRNA6		31	27	27.5	27.5			33	
RsdsRNA7			30	32	30			30	

RsdsRNA8		31	32	33			34	
	IBRS22	**IBRS23**	**IBRS04**	**IBRS07**	**IBRS11**	**IBRS15**	**IBRS16**	**IBRS19**
RsdsRNA10	30	29		30	29		29	28
RsBeLV1								22
RsBPLV1		20.48						
RsFLV1				30				31
RsALV1								28.5
RsALV2		30			30			30
RsALV3		29						
RsMV21		15.5						
RsMV22		18		18	16			30
RsMV23				14.5				
RsMV24	26	24	29	25	23	25.5	26	24
RsMV25			26					
RsMV26			25					
RsMV27			27					
RsMV28								21
RsMV29	24.5	22.5	28	22	21	21.5	22.5	22.5
RsMV30								21.3
RsMV31				14				
RsMV32		15						
RsMV33			25					
RsMV34			29					
RsMV35			27	29				17

	IBRS22	**IBRS23**	**IBRS04**	**IBRS07**	**IBRS11**	**IBRS15**	**IBRS16**	**IBRS19**
RsMV37		22	28	23	23			
RsMV38		17		16		34		
RsOLV2		20			16		26	21
RsOLV3				24				28
RsOLV4	30.5	25		25			30.5	27
RsOLV5				24.44				
RsFV1						26		
RsFV2	26.5	25		25.5	24		26	26.5
RsFV3		29	29	29				28
RsHV1							33	19
RsHV2							23	
RsHV3							24.55	
RsPuV1	23	22	28	21.5		21	22	
RsPuV2							24	20
RsPuV3	29	26	34	25.5			26.5	26
RsPuV4				22				

2.8 References

Andika, I.B., Wei, S., Cao, C., Salaipeth, L., Kondo, H., Sun, L. (2017). Phytopathogenic fungus hosts a plant virus: A naturally occurring cross-kingdom viral infection. Proc Natl Acad Sci USA **114**, 12267−12272.

Antoniolli, D. (2015). Produção, regularização e conquistas do mercado de gramas cultivadas no Brasil. In: Tópicos atuais em gramados IV, VII SIGRA – Simpósio sobre Gramados. FCA/UNESP. Mateus, C.M.D'A., Villas Bôas, R.L., Andrade, T.F., Oliveira, M.R., Backes, C., Santos, A.J.M., Godoy, L.J.G. (eds) Editora Fepaf, Botucatu, SP, Brazil, pp 9−22.

Bartholomäus, A., Wibberg, D., Winkler, A., Pühler, A., Schlüter, A., Varrelmann, M. (2017). Identification of a novel mycovirus isolated from Rhizoctonia solani (AG 2-2 IV) provides further information about genome plasticity within the order Tymovirales. Arch. Virol. **162**, 555-559.

Bharathan, N., Saso, H., Gudipati, L., Bharathan, S., Whited, K. (2005). Double-stranded RNA: distribution and analysis among isolates of Rhizoctonia solani AG-2 to -13. Plant Pathology **54**, 196-203.

Chiba, S., Salaipeth, L., Lin, Y.H. (2009). A novel bipartite double-stranded RNA mycovirus from the white root rot fungus Rosellinia necatrix: molecular and biological characterization, taxonomic considerations, and potential for biological control. J. Virol. **3**, 12801–12812.

Chu, Y.M., Jeon, J.J., Yea, S.J., Kim, Y.H., Yun, S.H., Lee, Y.W., Kim, K.H. (2002). Double-stranded RNA mycovirus from Fusarium graminearum. Appl. Environ. Microbiol. **68**, 2529-2534.

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**.

Das, S., Fallon, R.E., Stewart, A., Pitman, A.R. (2016). Novel mitoviruses in Rhizoctonia solani AG-3PT infecting potato. Fungal Biol. **120**, 338-350.

Diep Thi, H., Chernomor, O., von Haeseler, A., Minh, B. Q., Le Sy, V. (2018). UFBoot2: Improving the Ultrafast Bootstrap Approximation. Molecular Biology and Evolution **35**, 518-522.

Donaire, L., and Ayllon, M. A. (2017) Deep sequencing of mycovirus-derived small RNAs from Botrytis species. Molecular Plant Pathology, **18**, 1127-1137.

Drinnenberg, I. A., Fink, G. R., Bartel, D. P. (2011). Compatibility with killer explains the rise of RNAi-deficient fungi. Science **333**, 1592-1592.

Goic, B., Stapleford, K. A., Frangeul, L., Doucet, A. J., Gausson, V., Blanc, H., Schemmel-Jofre, N., Cristofari, G., Lambrechts, L., Vignuzzi, M., Saleh, M.-C. (2016). Virus-derived DNA drives mosquito vector tolerance to arboviral infection. Nature Communications, **7**.

Haas, B.J., Papanicolaou, A., Yassour, M., Grabherr, M., Blood, P.D., Bowden, J., Couger, M.B., Eccles, D., Li, B., Lieber, M., MacManes, M.D., Ott, M., Orvis, J., Pochet, N., Strozzi, F., Weeks, N., Westerman, R., William, T., Dewey, C. N., Henschel, R., LeDuc, R. D., Friedman, N., Regev, A. (2013). De novo transcript sequence reconstruction from RNA-Seq: reference generation and analysis with Trinity. Nature Protocols **8**, 1094–1515.

Hahn, C., L. Bachmann & B. Chevreux (2013) Reconstructing mitochondrial genomes directly from genomic next-generation sequencing reads-a baiting and iterative mapping approach. Nucleic Acids Research, **41**.

Hong, Y.G., Dover, S.L., Cole, T.E., Brasier, C.M., Buck, K.W. (1999). Multiple mitochondrial viruses in an isolate of the Dutch elm disease fungus Ophiostoma novo-ulmi. Virology **258**, 118–127.

Huang, X. Q., and Madan, A. (1999). CAP3: A DNA sequence assembly program. Genome Research **9**, 868-877.

Kalyaanamoorthy, S., Bui Quang, M., Wong, T. K. F., von Haeseler, A., Jermiin, L. S. (2017). ModelFinder: fast model selection for accurate phylogenetic estimates. Nature Methods **14**, 587-589.

Lakshman, D. K., Jian, J. H., Tavantzis, S. M. (1998). A double-stranded RNA element from a hypovirulent strain of Rhizoctonia solani occurs in DNA form and is genetically related to the pentafunctional AROM protein of the shikimate pathway. Proceedings of the National Academy of Sciences of the United States of America, **95**, 6425-6429.

Lam-Tung, N., Schmidt, H. A., von Haeseler, A., Bui Quang, M. (2015). IQ-TREE: A Fast and Effective Stochastic Algorithm for Estimating Maximum-Likelihood Phylogenies. Molecular Biology and Evolution **32**, 268-274.

Li, H., and Durbin, R. (2009) Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics **25**, 1754-1760.

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R. (2009). Genome Project Data The Sequence Alignment/Map format and SAMtools. Bioinformatics, **25**, 2078-2079.

Liu, L., Xie, J., Cheng, J., Fu, Y., Li, G., Yi, X., Jiang, D. (2014). Fungal negative-stranded RNA virus that is related to bornaviruses and nyaviruses. Proceedings of the National Academy of Sciences of the United States of America **111**, 12205-12210.

Marchler-Bauer, A., Bo, Y., Han, L., He, J., Lanczycki, C. J., Lu, S., Chitsaz, F., Derbyshire, M. K., Geer, R. C., Gonzales, N. R., Gwadz, M., Hurwitz, D. I., Lu, F., Marchler, G. H., Song, J. S., Thanki, N., Wang, Z., Yamashita, R. A., Zhang, D., Zheng, C., Geer, L. Y., Bryant, S. H. (2017). CDD/SPARCLE: functional classification of proteins via subfamily domain architectures. Nucleic Acids Research **45**, D200-D203.

Marquez, L. M., Redman, R. S., Rodriguez, R. J., Roossinck, M. J. (2007). A virus in a fungus in a plant: Three-way symbiosis required for thermal tolerance. Science **315**, 513-515.

Marzano, S.-Y. L., and Domier, L. L. (2016). Novel mycoviruses discovered from metatranscriptomics survey of soybean phyllosphere phytobiomes. Virus research **213**, 332-342.

Marzano, S.-Y. L., B. D. Nelson, O. Ajayi-Oyetunde, C. A. Bradley, T. J. Hughes, G. L. Hartman, D. M. Eastburn, Domier, L. L. (2016). Identification of Diverse Mycoviruses through Metatranscriptomics Characterization of the Viromes of Five Major Fungal Plant Pathogens. Journal of Virology **90**, 6846-6863.

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

Nerva, L., A. Silvestri, M. Ciuffo, S. Palmano, G. C. Varese, Turina, M. (2017). Transmission of Penicillium aurantiogriseum partiti-like virus 1 to a new fungal host (Cryphonectria parasitica) confers higher resistance to salinity and reveals adaptive genomic changes. Environmental microbiology **19**, 4480-4492.

Nerva, L., Varese, G. C., Turina, M. (2018b). Different Approaches to Discover Mycovirus Associated to Marine Organisms. Methods in molecular biology (Clifton, N.J.) **1746**, 97-114.

Nerva, L., G. Vigani, D. Di Silvestre, M. Ciuffo, M. Forgia, W. Chitarra, Turina, M. (2019). Biological and molecular characterization of Chenopodium quinoa mitovirus 1 reveals a distinct sRNA response compared to cytoplasmic RNA viruses. Journal of virology. JVI.01998-18.

Nerva, L., M. Ciuffo, M. Vallino, P. Margaria, G. C. Varese, G. Gnavi, and Turina, M. (2016). Multiple approaches for the detection and characterization of viral and plasmid symbionts from a collection of marine fungi. Virus Research **219**, 22-38.

Nerva, L., W. Chitarra, I. Siciliano, F. Gaiotti, M. Ciuffo, M. Forgia, Varese G. C., Turina, M. (2018a). Mycoviruses mediate mycotoxin regulation in Aspergillus ochraceus. Environmental microbiology.

Nibert, M. L., M. Vong, K. K. Fugate, Debat, H. J. (2018). Evidence for contemporary plant mitoviruses. Virology **518**, 14-24.

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

Nuss, D.L. (1992). Biological control of chestnut blight: an example of virus-mediated attenuation of fungal pathogenesis. Microbiol. **56**, 561-576.

Petrzik, K., Sarkisova, T., Starý, J., Koloniuk, I., Hrabáková, L., Kubešová, O. (2016). Molecular characterization of a new monopartite dsRNA mycovirus from mycorrhizal Thelephora terrestris (Ehrh.) and its detection in soil oribatid mites (Acari: Oribatida). Virology **489**, 12-19.

Picarelli, M.A.S.C. (2015). Estudo de micovirus em Rhizoctonia solani como estratégia para controle biológico de rizoctoniose em gramados. [dissertation]. [São Paulo (SP)]: Instituto Biológico, Brazil. Available at: http://www.biologico.agricultura.sp.gov.br/pos/uploads/files/pdf/2015/aurea.pdf

Shi, M., Lin, X.-D., Tian, J.-H., Chen, L.-J., Chen, X., Li, C.-X., Qin, X.-C., Li, J., Cao, J.-P., Eden, J.-S. (2016). Redefining the invertebrate RNA virosphere. Nature **540**, 539-543.

Sigrist, C. J. A., Castro, E. de, Cerutti, L., Cuche, B. A., Hulo, N., Bridge, A., Bougueleret, L., Xenarios, I. (2013). New and continuing developments at PROSITE. Nucleic Acids Research **41**, E344-E347.

Strauss, E. E., Lakshman, D. K., Tavantzis, S. M. (2000). Molecular characterization of the genome of a partitivirus from the basidiomycete Rhizoctonia solani. Journal of General Virology **81**, 549-555.

Tamura, K., Stecher, G., Peterson, D., Filipski, A., Kumar, S. (2013). MEGA6: Molecular Evolutionary Genetics Analysis Version 6.0. Molecular Biology and Evolution **30**, 2725-2729.

Trifinopoulos, J., Lam-Tung, N., von Haeseler, A., Minh, B. Q. (2016). W-IQ-TREE: a fast online phylogenetic tool for maximum likelihood analysis. Nucleic Acids Research **44**, W232-W235.

Turina, M., Ghignone, S., Astolfi, N., Silvestri, A., Bonfante, P., Lanfranco, L. (2018). The virome of the arbuscular mycorrhizal fungus Gigaspora margarita reveals the first report of DNA fragments corresponding to replicating non-retroviral RNA viruses in Fungi. Environmental microbiology.

Untergasser, A., I. Cutcutache, T. Koressaar, J. Ye, B. C. Faircloth, m. Remm & S. G. Rozen (2012) Primer3-new capabilities and interface. Nucleic Acid Research, **40**.

Vainio, E. J., Jurvansuu, J., Streng, J., Rajamaki, M. L., Hantula, J., Valkonen, J. P. T. (2015). Diagnosis and discovery of fungal viruses using deep sequencing of small RNAs. Journal of General Virology **96**, 714-725.

Wolf, Y. I., Kazlauskas, D., Iranzo, J., Lucia-Sanz, A., Kuhn, J. H., Krupovic, M., Dolja, V. V., Koonin, E. V. (2018). Origins and Evolution of the Global RNA Virome. Mbio **9**.

Wu, M.D., Zhang, L., Li, G.Q., Jiang, D.H., Hou, M.S., Huang, H.C. (2007). Hypovirulence and double-stranded RNA in Botrytis cinerea. Phytopathology **97**, 1590-1599.

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. Proceedings of the National Academy of Sciences of the United States of America **107**, 8387-8392..

Zanon, M.E. (2015). Desenvolvimento de grama 'Esmeralda', grama bermudas 'Tifway 410' e 'Celebration' submetidas a aplicação de reguladores de crescimento. Ph.D. thesis Jaboticabal(SP): Universidade Estadual Paulista/UNESP, Jaboticabal.

Zheng, L., Liu, C., Zhang, M., Yang, M., Zhou, E. (2018). Diversity of dsRNA viruses intecting rice sheat blight. Rice Science **25**, 57-60.

Zheng, L., Zhang, M., Chen, Q., Zhu, M., Zhou, E. (2014). A novel mycovirus closely related to viruses in the genus Alphapartitivirus confers hypovirulence in the phytopathogenic fungus Rhizoctonia solani. Virology **456/457**, 220-226.

Zhong, J., Chen, C.Y., Gao, B.D. (2015). Genome sequence of a novel mycovirus of Rhizoctonia solani, a plant pathogenic fungus. Virus Gene **51**, 167-170.

Chapter 3

Virome characterization of *Cryphonectria parasitica* isolates from Azerbaijan unveiled a new mymonavirus and a putative new RNA virus unrelated to described viral sequences

Marco Forgia[1,2], Elnare Isgandarli[3], Dilzara N. Aghayeva[3], Irada Huseynova[4], Massimo Turina[1]

1 Institute for Sustainable Plant Protection, CNR, Strada delle Cacce 73, 10135, Torino, Italy
2 Department of Life Sciences and Systems Biology, University of Turin, Viale Mattioli 25, 10125, Torino, Italy
3 Institute of Botany, Azerbaijan National Academy of Sciences, Badamdar highway 40, Baku, AZ1004, Azerbaijan
4 Institute of Molecular Biology and Biotechnologies, Azerbaijan National Academy of Sciences, Matbuat avenue 2a, Baku, AZ1073, Azerbaijan

Published on Virology
DOI: https://doi.org/10.1016/j.virol.2020.10.008

3.1 Abstract

Cryphonectria parasitica, the causal agent of chestnut blight is controlled in Europe through natural spread of Cryphonectria hypovirus 1 (CHV1), a mycovirus able to induce hypovirulence to the host. In recent years *C. parasitica* was reported infecting Azerbaijani population of chestnut, but the presence of CHV1 still needs to be confirmed. Aim of this work was to investigate fifty-five *C. parasitica* isolates collected in Azerbaijan to describe the associated viruses. Our work found i) the first negative-sense ssRNA virus known to infect *C. parasitica* naturally for which we propose the name Cryphonectria parasitica sclerotimonavirus 1 (CpSV1) and ii) an RNA sequence showing peculiar features suggesting a viral nature for which we propose the name Cryphonectria parasitica ambivirus 1 (CpaV1). The discovery of CpaV1 expands our knowledge of the RNA virosphere suggesting the existence of a new lineage that cannot presently be reliably associated to the monophyletic *Riboviria*.

Keywords

Mycovirus; fungi; mymonaviridae; ORFan; chestnut; Cryphonectria parasitica;

3.2 Introduction

Cryphonectria parasitica is the ascomycete causal agent of chestnut blight disease that endangered chestnut cultivation in Europe and America during the 20th century (Dawe and Nuss, 2013). In Europe, chestnut blight disease control was achieved through spreading of the Cryphonectria hypovirus 1 (CHV1) that causes hypovirulence in the fungal host (Grente, 1965; Milgroom and Cortesi, 2004; Turina and Rostagno, 2007; Rigling and Prospero, 2018). After the setup of efficient reverse genetic systems based on different CHV1 infectious clones and the availability of protocols for protoplast production and genetic transformation, *C. parasitica* became an important model to study hypovirulence and host-mycovirus interaction (Nuss, 2005). Nevertheless, a relatively low number of mycovirus infecting *C. parasitica* have been described so far. Indeed, even if Next Generation Sequencing (NGS) approaches have been shown to represent the most valuable tool to deeply investigate the viral diversity in a collection of fungal isolates (Nerva *et al.*, 2016; Marzano *et al.*, 2016; Donaire and Ayllon, 2017), to our knowledge there are no works in literature implementing these approaches on *C. parasitica,* and most of the mycoviruses infecting this model fungus have been identified through dsRNA isolation and traditional cloning and sequencing of fragments of cDNA (Hillman and Suzuki, 2004). The mycoviruses infecting *C. parasitica* characterized so far show different genome organizations and have dsRNA genomes or positive ssRNA genomes. Taxonomically they belong to different families, *Hypoviridae, Reoviridae* and *Mitoviridae,* but the best studied among them is the hypovirus CHV1; so far, no negative stranded mycovirus has been isolated from *C. parasitica* (Eusebio-Cope *et al.*, 2015). Apart from naturally occurring infections, different works showed the possibility to infect *C. parasitica* with different mycoviruses through purified viral particle transfection allowing host-virus interaction studies even with viruses not originally isolated from the

model fungal host (Nerva *et al.*, 2017; Salaipeth *et al.*, 2014); conversely, CHV1 was shown to potentially infect other fungi, where it facilitates plant virus co-infection (Bian *et al.*, 2020).

During recent years *C. parasitica* has been reported in Azerbaijan where chestnut cultivation occurs (Aghayeva and Harrington, 2007). *C. parasitica* isolates collected in Azerbaijan show a high genetic homology compared to those from neighbor Georgia. In facts, the majority of Azerbaijani vegetative compatibility types are present in Georgia and belong to the Georgian gene pool of *C. parasitica* (Aghayeva et al., 2017; Prospero et al., 2013). To date, little is known about the dissemination of CHV1 through Azerbaijani population of *C. parasitica*; aim of this work was to collect a large number of fungal isolates from Azerbaijani population of *C. parasitica* and describe through a high-throughput approach the virome of the collected isolates in order to find evidence of CHV1 (or other hypovirus) infections and to describe possible new mycoviruses infecting this fungal host. CHV1 detection in Azerbaijani isolates could give information about the possibility to control chestnut blight disease by spreading hypovirulent isolates; furthermore, the discovery of new viruses infecting *C. parasitica* as natural host could allow investigating host-virus interaction at the molecular level thanks to the possibilities of multiple forward and reverse genetic approaches allowed by this genetically treatable fungal model.

3.3 Results

3.3.1 Isolates collection and identification

Isolation of *C. parasitica* was performed in three districts (Qabala, Shaki, Zagatala; Fig. 1). Bark samples were collected from a total of 77 cankers and from 58 of them *C. parasitica* was isolated in pure culture. Most of the remaining bark samples also yielded *C. parasitica* colonies, but these were heavily contaminated with other fungal species and were not included in further analyses. Some variation was observed among the isolates with respect to their culture morphology (i.e., pigmentation and asexual sporulation) on PDA. Thirty-seven cultures showed an orange phenotype, while 21 displayed a whitish-orange phenotype on PDA 10 days post inoculation. White cultures, typical of *C. parasitica*-infected with CHV-1 were not observed. Incidence of sexual reproduction was estimated observing *C. parasitica* stromata under the microscope. Both perithecia and pycnidia were observed in most examined stromata. All examined perithecia contained asci with ascospores. Perithecia were present in all districts with the highest prevalence found in Zagatala (56%). Both mating types of the fungus were found in all three districts. From the isolates collected, fifty-five were chosen for virome exploration through NGS approach. A list of isolates with information about collection site and date can be found in Supp. Table S1.

3.3.2 Virome exploration

Fifty-five isolates from the Azerbaijani *C. parasitica* collection were inoculated in liquid cultures to obtain abundant mycelia. From the obtained mycelia, total RNAs were extracted and pooled in two samples for Illumina sequencing; pooled samples composition is showed in Supp. Table S2. After ribosomal RNA depletion, library

construction and Illumina sequencing, the reads obtained from the two libraries were assembled and viral contig were retrieved through a blast-based bioinformatic pipeline. From this analysis, we were able to identify one virus related to the *Mymonaviridae* family named Cryphonectria parasitica sclerotimonavirus 1 (CpSV1) and one ORFan sequence with peculiar features suggesting a viral origin. Information about read abundance in the NGS sample hosting the viruses are shown in Table 1. Our virome characterization could not provide evidence of hypovirus infection, and specifically, CHV1 could not be detected, either by NGS analysis or by specific qRT-PCR (not shown).

3.3.3 CpSV1 characterization

After blasting the transcriptomes against our custom viral database, we identified one 6439 bp long contig coding for three major ORFs in NGS sample 1 named Contig1. The third ORF was partial and showed a *Mononegavirales* RdRP domain with a close homology to RdRP encoded by the sclerotimonavirus Soybean leaf-associated negative-stranded RNA virus 2. Quantitative RT-PCR analysis on cDNAs performed on each of the fungi included in the NGS Sample 1 revealed that Contig1 was present only in the isolate ACP 28 (Supp. Table S2) and no amplification was observed using the same qRT-PCR primer on ACP 28 DNA (data not shown); this result confirmed the viral nature of Contig1 excluding the possibility that the virus is integrated in the fungal genome. Viruses belonging to *Mymonaviridae* family have a negative single-stranded RNA genome and encode for an ORF in the 5' extremity downstream the RdRP-encoding ORF, which is missing in our contig. To complete the sequence of the *C. parasitica* viral contig, we retrieved the 5' terminal ORF from Soybean leaf-associated negative-stranded RNA virus 2 (one of the viruses showing higher homology to our contig) and we blasted it against our transcriptome from NGS Sample 1. We identify one 2901 bp contig named Contig2 hosting the remaining part of the genome encoding the mymonavirus RdRP and one ORF in the 5' part of the contig. Since Contig1 and Contig2 did not overlap, we designed a primer pair to amplify the putative missing part to complete the RdRP coding sequence. After RT-PCR amplification, cloning of the PCR product and sequencing of 6 different clones we obtained the consensus sequence of the missing part of the viral genome between Contig1 and Contig2 and we named the full length virus Cryphonectria parasitica sclerotimonavirus 1 (CpSV1). Rapid Amplification of cDNA ends (RACE) on the 5' and 3' extremity of the viral genome were performed and results added just two nucleotides at the 5' end and one nucleotide at the 3' end (not shown). Furthermore, in silico assembly of the genome was confirmed by RT-PCR amplification of 1 kb overlapping fragments covering most of the viral sequence (Suppl. Fig. S1a; the primers used are listed in Suppl. Table S3**)**. Taken together these results show the genome organization and the main features of the first complete sequence of a negative-sense single stranded RNA virus infecting *C. parasitica* (Fig. 2). CpSV1 is 9318 bp long and it encodes five ORFs named ORF 1 to 5. Blastp results obtained submitting each ORF against the NCBI database are shown in Table 2. ORF1 produces a 293 amino acid long putative protein that shows an USP19_linker domain when submitted to NCBI domain finder and the blastp analysis shows little homology with other putative proteins encoded by mycoviruses belonging to *Mymonaviridae* family. ORF2 codes for what is likely the nucleoprotein of CpSV1, a predicted 391 amino acid long protein showing close homology to other putative

nucleoproteins encoded by mymonaviruses; the predicted protein does not show any conserved domain on NCBI domain finder. ORF3 is a 42 amino acid protein without homology to proteins in databases. ORF4 encodes for the viral RdRP that is predicted to be 1931 amino acid long and shows a *Mononegavirales* RdRP domain together with a *Mononegavirales* mRNA capping domain. The last putative protein, produced by ORF5, is 178 amino acid long and displays homology with proteins encoded from the 5' extremity of several mycoviruses belonging to the *Mymonaviridae* family; no conserved domains are detected searching with NCBI domain finder. A typical feature showed by mymonaviruses is the conservation of the sequences in between the ORFs and between the last ORF and 5'UTR extremity. Indeed, by aligning these sequences retrieved from CpSV1 we were able to observe a conserved nucleotide pattern that is shown in Fig. 2b. Phylogenetic analysis performed on the RdRP sequences (Fig. 3) placed CpSV1 in the *Mononegavirales* order and in the *Mymonaviridae* family. CpSV1 is likely a member of the only genus so far described in the *Mymonaviridae* family called *Sclerotimonavirus* that hosts only mycoviruses; viruses included in this genus are highlighted in the tree and are found in the same monophyletic branch that is statistically well supported by bootstrap analysis.

3.3.4 Characterization of an ORFan sequence as a putative viral segment

Since a major limit of homology-based methods to find viruses is the loss of viral contigs with low or no homology to known sequences present in the database, we looked into the possibility of finding new viral sequences in ORFans contig. For this purpose, we filtered the transcriptomes to remove the fungal transcript and retrieve only the contigs showing "no hit" to any sequence contained in the *C. parasitica* genomic protein collection. The same procedure was performed on a pooled transcriptome we used to describe *Rhizoctonia solani* virome in a previous work (Picarelli *et al.*, 2019), against *R. solani* proteome. After this analysis, we removed the smallest transcripts and selected contigs coding for proteins with a molecular mass of at least 15 kDa. These putative proteins, retrieved from *C. parasitica* and *R. solani* RNAseq performing the same pipeline, were blasted against each other. Homologous proteins were found to be encoded by an ORFan contig of 4623 bp coding for 2 putative ORFs, present in both *C. parasitica* NGS sample 1 and 2, and two ORFan contigs of 4868 bp and 3729 bp present in *R. solani* NGS sample. Based on the ambisense nature of the putative viral contigs, the three contigs were provisionally named Cryphonectria parasitica ambivirus 1 (CpaV1), and Rhizoctonia solani ambivirus 1 and 2 (RsaV1 and 2). A schematic representation of the genome organization of CpaV1 with typical ambisense orientation of two ORFs can be observed in Fig. 4. These contigs show similar ORFs pattern and conserved domains among homologous proteins (Fig. 4b and 4c) and since in all cases one of the two predicted proteins contains a GDD triplet, the hallmark of RNA-dependent RNA-polymerase, further *in vivo* characterizations of these genome organization have been performed.

To assess CpaV1 prevalence in the fungal collection we performed qRT-PCR analysis on all the *C. parasitica* isolates cDNA (primers in Suppl. Table S3**)** and we found (Suppl. Table S2**)** that twenty-two out of fifty-five isolates host CpaV1. Moreover, since the prevalence was high in the Azerbaijani isolates, we checked seven Italian isolates from North-western Alps that we had in our collection and we found them all CpaV1-free (Suppl. Table S4). For

further characterization we used as a reference isolates ACP 34, 36 and 43**.** Since no signal was detected in q-PCR reactions performed on DNA extracted by infected *C. parasitica,* we can assume that CpaV1 is an RNA sequence that does not derive from transcription of cDNA integrated in the fungal genome (Crouch *et al.*, 2020) nor from transcription of a DNA virus. We can also exclude that these putative viral sequences are replicated via a DNA intermediate. Further confirmation of this evidence has been obtained by PCR amplification of three hundred bp long sequences from both ORFA and ORFB coding sequences (Suppl. Fig. S2). Indeed, the amplicon was obtained only in the sample containing cDNA as a template and no amplification was observed in the samples containing ACP 34, 36 and 43 DNA. PCR fragments obtained from this analysis were cloned and sequenced to use them as probes for northern blots (see below). Attempts to purify dsRNA (a typical signature of some mycoviruses) from CpaV1 infected isolates were made on isolate ACP 43 using isolate 54 as negative control, but no evidence of dsRNA accumulation was observed on ethidium bromide stained electrophoresis gel (data not shown).

Remarkably even if the two ORFs showed conservation compared to contigs isolated from different fungi, they had different orientation: the two ORFs are always disposed in an ambisense orientation, but they could be oriented outward (pointing to the 5' and 3' UTRs as observed in *C. parasitica* ORFan contig) or inward, toward the center of the contig (as observed in the two contigs from *R. solani*). This evidence brought to the hypothesis that these contigs could derive from a circular sequence. We thus designed PCR primers amplifying upstream the 5' end and downstream the 3' end, across a putative 3'-5' covalent link, and we obtained a RT-PCR product that was cloned and sequenced, indeed showing the possibility to reconstruct a circular sequence or a dimer sequence from the initial contig and allowing us to confirm that the sequence is complete. To further confirm that CpaV1 sequence is complete, we amplified 1kbp long overlapping PCR fragments covering the entire sequence and results confirmed the expected size of each portion (Suppl. Fig. S1b). As shown in Fig. 4 the two ORFs (named ORFA and ORFB) have an ambisense orientation on the genome fragment. ORFA is a 757 amino acid long protein and it doesn't show any conserved domain; blastp analysis (Table 2) shows homology with a partial protein encoded by an ORFan sequence isolated from *Agaricus bisporus* and that belongs to an RNA sequence not integrated in the genome (Deakin *et al.*, 2017). The predicted protein hosts a GDD triplet that is fundamental for subdomain C of the viral RdRps (te Velthuis, 2014). ORFB encodes for a 663 amino acid sequence that has homology (Table 2) with hypothetical proteins from several fungi; these proteins could derive from integration events happened between host fungi and other CpaV1-like sequences. Northern blot analysis were performed using fragments amplified on both ORFs in each sense to detect positive and negative sense RNA (considering as positive sense the one hosting ORFA in the positive orientation); analysis was carried out using ACP 34 and 36 as positive isolates and ACP 23 as negative (see Fig. 4a, for information about northern blot probes position and orientation on the sequence). Results showed two bands specific to the infected isolates that are present with all the probe tested directed to ORFA or ORFB (Fig. 5); short exposure time allows to appreciate that the smallest of the two bands appear to be less intense than the largest

one and probes detecting the putative negative sense RNA give a less intense signal for all ORFs. ACP 23 isolate included as negative control showed few bands that are unspecific signal coming from probe-ribosomal RNA interaction (confirmed overlapping with methylene blue stained membrane). Comparing these bands to the ones obtained by hybridization with a Tomato brown rugose fruit virus (TBRFV) infected tomato (Salem *et al.*, 2016) that was added on the RNA gel as a size-marker, we could estimate that the lower band is compatible with CpaV1 linear genomic size and we hypothesize that the higher band could derive from a structure assumed by the circular entity of the genome causing a shift in the electrophoresis migration or from a possible linear dimer. To understand if CpaV1 really has a circular sequence we treated extracted RNA from ACP 34 with RNAse R enzyme, that is an exonuclease working specifically on linear RNAs and circular RNA molecules should be naturally protected from his action. RNA from ACP 34 and TBRFV infected tomato were treated with RNAse R for thirty minutes, one hrs and two hrs respectively, then the RNAs were migrated and blotted to perform northern blot analysis. Supplementary on line Fig. S3 shows results obtained from this experiment, where RNAse R action was confirmed because ribosomal RNAs were not visible anymore even after only thirty minutes treatment but they were still present in the samples just mixed with RNAse R reaction buffer. Both CpaV1 and TBRFV, used as negative control, didn't give any signal after hybridization with specific northern blot probes suggesting that CpaV1 is not circular. This result also suggests that the higher band detected in northern blots is probably a dimer formed by two CpaV1 molecules connected head to tail (explaining the possibility to amplify RT-PCR segments across the putative junction). Looking at northern blot results, it appears that the most abundant form of CpaV1 is the one where ORFA is found in positive orientation, with the higher band (probably associated to the dimer form) more intense than the lower one, suggesting that the putative plus strand dimer is likely the most abundant virus RNA form in the host. To further investigate if protein encoded by ORFA (carrying the GDD triad) could be the viral replicase, we retrieved from a number of characterized viral RdRp submotifs A, B and C from the palm domain (Gorbalenya *et al.*, 2002). By aligning these conserved regions to ORFAs-encoded proteins from CpaV1, RsaV1 and RsaV2 we could observe a partial conservation of the key amino acid residues. Indeed, the key D residue from motif A and the GDD triad from motif C could be found in the ambivirus proteins, while the conserved residues from motif B are not identifiable (Suppl. Fig. S4).

Forty-eight single conidia cultures were obtained from CpaV1 infected isolate ACP 34 to possibly obtain isogenic strains infected and non-infected with CpaV1 and characterize possible hidden effects of the ambivirus on the host. All the colonies obtained resulted non-infected when testing them through qRT-PCR, suggesting that CpaV1 is easily lost through conidia propagation (at least in laboratory condition) (not shown).

3.4 Discussion

Cryphonectria hypovirus 1 can be found across Europe, in Asia and in some parts of North America. In the latter area the dissemination was performed artificially for biological control, resulting less efficient than in other countries likely because of a higher genetic diversity of *C. parasitica* population (Rigling and Prospero, 2018) that would not allow efficient CHV1 transmission among vegetatively incompatible isolates. CHV1 was also detected in Georgia, which is likely the source of Azerbaijani *C. parasitica* isolates (Rigling *et al.*, 2018). After CHV1, three other hypoviruses have been reported to infect *C. parasitica*: Cryphonectria hypovirus 2, 3 and 4. All these hypoviruses are able to induce hypovirulence, even if most of the works are concentrated on CHV1 because of the easy manipulation and its better performance as biocontrol agent. CHV2, 3 and 4 have all been found in North America, but CHV2 was also detected in China; in both cases, the areas where *C. parasitica* is infected by these viruses are limited (Hillman and Suzuki, 2004). Using high throughput approaches we demonstrated that our collection of *C. parasitica* isolates from Azerbaijan doesn't host Cryphonectria hypovirus 1 (or any of the other hypoviruses found in other populations) and that spreading of the hypovirulent strain can not rely on a CHV1-infected isolate already present in the environment. Moreover, this work unveiled Cryphonectria parasitica sclerotimonavirus 1, the first negative-sense single stranded RNA virus infecting *C. parasitica* as a natural host. This result could be important to characterize the interaction between these viruses and the host thanks to the possibility of genetic manipulations allowed by the fungal model. Indeed in some cases viruses belonging to the *Mymonaviridae* family and the *Sclerotimonavirus* genus have been associated to hypovirulence (Liu *et al.*, 2014), so further characterization could be made trying to obtain isogenic isolates not infected with CpSV1 to observe possible differences in pathogenicity.

The major limit of homology-based method to find viruses is the loss of viral contig with low or no homology to known sequences. Recently, two putative nucleocapsids from Bunya-like mycoviruses have been found, thanks to distant homology between the sequences, by comparing the contigs with no-hits in the NCBI database from two NGS experiments containing two different Bunya-like mycoviruses (Nerva *et al.*, 2019). We decided to perform a similar approach on *C. parasitica* NGS transcriptome comparing it to the transcriptome obtained from *Rhizoctonia solani* RNAseq from a previous work that contained more than sixty viruses (Picarelli *et al.*, 2019). Results allowed us to identify homologous ORFan contigs coding for two ORFs in ambisense orientation. The sequence found in *C. parasitica* belongs to an RNA molecule that is not integrated in the host genome, therefore supporting its viral origin (self-replicating RNA molecule). The possibility that another RdRP replicates in trans this molecule has been excluded by our screen of *C. parasitica* virome in which no other viral RdRPs with a conserved palm domain detectable by blast search, have been found in the samples. The presence of a GDD carrying subdomain in the amino acid sequence from putative ORFA can suggest that it is part of the RdRP palm domain and that the ORFan contig we detected could be a virus; furthermore, northern blot analysis together with RT-PCR analysis showed that a head to tail abundant dimer accumulation can be possible for CpaV1. CpaV1 has

homology with contigs isolated from a different fungal host, but ORF orientation is not conserved between these sequences, possibly because the ability to form dimer could lead to mistakes in the bioinformatic reconstruction of the ambivirus ORFan genome and its graphical representation. Abundant dimer accumulation during replication of negative strand RNA virus was previously reported (Bertran *et al.*, 2016) and hypothesized to be linked to a regulation of replication analogous to that caused by accumulation of defective interfering RNA (DI-RNA).

Notably, this is not the first time that ORFan sequences like CpaV1, RsaV1 and 2 are identified while exploring the virome of a fungus. During the characterization of the virome from a collection of isolates of *Agaricus bisporus,* several ORFan proteins belonging to RNA sequences not integrated in the fungal genome were detected thanks to a specific bioinformatic approach used for that study (Deakin et al., 2017). Among these, ORFan1 displays an organization that is possible to connect to a dimer version of a CpaV1-like sequence with a mis-assembly that leaves ORFA uncomplete. Indeed, comparing the four predicted proteins from ORFan1 with the CpaV1 putative proteins, the second and fourth proteins have the same sequence and share conservation with CpaV1 ORFB while the first and third proteins share only the first amino acids of their sequence (suggesting a bioinformatic error when assembling the dimeric version of ORFan1). Nevertheless, partial sequence of ORFan1 ORFA putative protein (at the 5' extremity of the contig) shows conserved motifs when aligned to ORFA from CpaV1 (Supp. Fig. S5). This is another evidence suggesting the possible viral nature of these sequences that have been found in several different fungal species.

It is interesting to point out the relatively high prevalence (twenty-two infected isolates out of fifty-five) of CpaV1 in the collection from Azerbaijan, particularly since we recently checked both by NGS and by qRT-PCR their presence in 7 Italian isolates from Cuneo and Torino province and we failed to detect it (Suppl. Table 4). Also, given the presence of fragments of those putative viruses in genomes of other fungi and their presence in at least two different species of Basidiomycota, *Rhizoctonia solani*, and *Agaricus bisporus*, we can expect that they could be widespread in fungi. We could not observe any macroscopic difference among CpaV1-infected and CpaV1-free isolates from the collection, even if we did not specifically test difference in virulence on chestnut in controlled environment or other more subtle features. This apparent symptomless infection, although not surprising for mycoviruses, could change its outcome in the case of interaction with CHV1, a virus that carries a suppressor of antiviral silencing, that can facilitate and exacerbate co-infections (Bian *et al.*, 2020) and that could be introduced for biocontrol purposes in Azerbaijan. Moreover, the possibility of antagonism between the two viruses leading to an inhibition of the biocontrol efficacy of CHV1 in coinfection with CpaV1 could not be excluded, and this could represent a problem for *C. parasitica* management strategy.

Attempts at partially purifying virus-like particle in CpaV1 infected isolates failed (not shown), but occurrence of viral replicons that do not assemble in a proper virion is a common feature of many mycoviruses, such as botourmiaviruses (Ayllon *et al.*, 2020), narna-like viruses (Hillman and Cai, 2013), hypoviruses (Suzuki *et al.*, 2018), and the recently characterized Hadaka virus 1 (Sato *et al.*, 2020).

The characterization of ambivirus sequences adds a further layer of diversity to know RNA viruses in the *Riboviria*, and raises the question of their position in the phylogenetic tree of viral RdRP (Wolf et al., 2018): at the moment the alignment is too poor to provide a reliable topology, but future discovery of other viral sequence might help obtaining more conserved alignments and a robust tree topology.

3.5 Methods

3.5.1 Chestnut blight detection and sampling

A total of three chestnut growing sites in three districts in North-Eastern Azerbaijan were surveyed in December 2018 and April 2019 (Fig. 1). Chestnut trees were visually inspected for symptoms of *C. parasitica*, such as bark cankers on stems or branches, or wilting leaves on individual branches in the crown (Rigling and Prospero, 2017). Bark pieces (ca. 2 × 2 cm) with visible *C. parasitica* stromata were removed with a knife from the cankers. Before collecting each canker sample, the knife was dipped in 70% ethanol and flamed. To avoid the isolation of clones of the same genotype, only one single canker per tree was sampled (Milgroom and Lipari, 1995). In each district, 25–30 cankers were sampled.

3.5.2 Fungal isolation, recognition and maintenance

Bark samples were examined under a dissecting microscope for the presence of pycnidia or perithecia of the fungus (Prospero *et al.*, 2013). Perithecia were dissected from the stromata and crushed in a drop of water to verify presence of asci with ascospores. Stromatal tissue of *C. parasitica* was removed from the bark samples with a sterile needle under a dissecting microscope and placed on potato dextrose agar (PDA: 39 g/L, Difco™, Becton, MD, USA) for isolation. The PDA plates were incubated at room temperature, and one outgrowing culture (isolate) per bark sample was kept for further analyses. Fungal isolates were maintained on 60 mm diameter Petri dishes containing PDA media at 26 °C with a light/dark cycle of 12 hours. To obtain large amount of fungal biomass for molecular analysis, fungi were inoculated in 50 ml of liquid EP complete media (Puhalla and Anagnostakis, 1971). After 3 days, liquid cultures were harvested with a Buchner filter covered with miracloth (Merck, Darmstadt, Germany) in a vacuum pump to separate mycelia from the liquid media. Obtained mycelia was lyophilized and kept at -20 °C. Italian isolates (Suppl. Table 4) came from a collection already present at the IPSP-CNR institute in Turin.

3.5.3 RNA, DNA and dsRNA extraction

RNA extraction was performed using Spectrum plant total RNA kit (Sigma-Aldrich, St. Louis, MO, USA) starting from 30 mg of lyophilized mycelia as previously described (Nerva *et al.*, 2018). DNA was extracted starting from 50 mg of lyophilized mycelia in a 2 ml O-ring tube with 0.5 mm diameter glass beads. Mycelia was broken using a Fast prep-24 (MP Biomedicals, Irvine, CA, USA), adding 700 µl of 2X STE/SDS and 700 µl of phenol pH 8.0 and repeating the fast prep step. Tubes were then centrifuged for 3 minutes in a table centrifuge at maximum speed for phase separation and upper phase was collected in new 2 ml tubes. Obtained phases were washed twice with equal volume of 24:1 chloroform: isoamyl alcohol and upper phases were collected every time. Nucleic acids in the upper layer were then precipitated with 2 volumes of 100% ethanol, and 5% of the phase volume of 3 M sodium acetate pH 3.5 and keeping the tubes at -80 °C for 20 minutes. Tubes were centrifuged 3 minutes at max speed and supernatants were discarded, the pellets were washed with 70% ethanol, and resuspended in 50 µl of sterile water and quantified with a Nanodrop 2000 (Thermo-Fisher scientific, Waltham, MA, USA).

Obtained DNAs were diluted to 10 ng/µl for PCR amplification use. dsRNA extraction was performed as described by Nerva and coworkers (Nerva *et al.* 2016).

3.5.4 Illumina sequencing and bioinformatic analysis

Total RNAs from each isolates of the collection of *C. parasitica* were pooled in two samples maintaining an equal proportion of RNA coming from each isolate. Composition of the two pooled samples is shown in Suppl. Table S2**.** Samples were then sent to Macrogen (Seul, South Korea) for ribosomal RNA depletion, library construction and Illumina sequencing using TruSeq Stranded Total RNA LT Sample Prep Kit (Gold) kit (Illumina, San Diego, CA, USA) and producing 100 million reads of 100 bp length in paired end. Reads were cleaned using bbmap functions in order to eliminate short reads, low quality reads and artifacts following a protocol from JGI available at the following doi (dx.doi.org/10.17504/protocols.io.gydbxs6). Clean reads were assembled using Trinity version 2.3 (Haas *et al.*, 2013) and obtained transcriptome were blasted against a custom viral database (https://osf.io/c9x2p/) using DIAMOND blastx function version 0.9.21 (Buchfink *et al.*, 2015). Blast results were manually checked to retrieve the viral contigs and to further characterize them in vivo. Mapping reads on the viral genomes were obtained producing an alignment in SAM format with bowtie2 version 2.2.9 (Langmead and Salzberg, 2012); SAM file was converted in BAM format with SAMTOOLS (version 1.3.1) view and sorted with SAMTOOLS sort (Li *et al.*, 2009). Using SAMTOOLS index a sort.bam.bai file was generated, and result was displayed on Tablet (Milne *et al.*, 2016). ORF prediction was performed using ORF finder NCBI tool and conserved protein domains were searched through submission on NCBI conserved domain search. Bioinformatic pipeline adopted is well described in a previous publication (Nerva *et al.*, 2018). ORFan investigation was conducted through a modified version of the method used in previous publications (Nerva *et al.*, 2019; Chiapello *et al*., 2020): Trinity assembly from *C. parasitica* was cleaned from contigs encoding for genomic proteins through DIAMOND tblastn against the *C. parasitica* genomic protein database downloaded from JGI platform (https://genome.jgi.doe.gov/portal/Crypa2/Crypa2.download.html). Contig showing no hit were retrieved from the blast results and filtered to obtain sequences longer than 1 kbp using R libraries Tidyverse ans Biostrings. Obtained contigs were submitted to NCBI ORF finder to predict possible ORFs and the predicted protein showing the longest amino acid length was kept for each submission. Obtained proteins were filtered to obtain just the proteins above 150 amino acids long. The same procedure was performed on a pooled transcriptome containing different *R. solani* isolates used in previous works (Picarelli *et al.*, 2019); this time, after obtaining transcripts over 1 kbp not blasting with the proteins encoded by the genome (Wibberg *et al.*, 2013), we used these sequences to build a blast database (using blast version 2.2.31) and the tblastn algorithm to compare the proteins retrieved for *C. parasitica* with predicted protein from *R. solani* contigs with a evalue cutoff of 0.00001. Blast results was manually checked for convincing alignments and proteins were kept and submitted to NCBI for a blastp analysis against the NR database to exclude a recent characterization and confirm their ORFan nature.

3.5.5 qRT-PCR analysis

To associate each viral contig to the correct fungal isolate contained in the pooled sample, qRT-PCR analysis was performed after designing specific primers on each viral

contig using Primer3 (Untergasser *et al.*, 2012) allowing an amplification size between 75 and 120 nucleotides. qRT-PCR were performed on a CFX connect (Bio-rad, Hercules, CA, USA) apparatus and each reaction was prepared in 10 µl using the same protocol and reagents as described in previous publication (Picarelli *et al.*, 2019).

3.5.6 PCR analysis, cloning and northern blot probes preparation

RT-PCR was performed to obtain cDNAs from infected isolates to produce northern blot probes for CpaV1 and to obtain the missing part of the viral genome for Cryphonectria parasitica sclerotimonavirus 1. cDNAs were synthesized from the RNA extracts using high capacity cDNA kit (Thermo-Fisher scientific, Waltham, MA, USA), following manufacturer instructions and obtained cDNAs were diluted 1:5 in sterile water. RT-PCRs were performed using One Taq DNA polymerase (NEB, Ipswich, MA, USA) following manufacturer instruction; bands were isolated from agarose gel using Zymo DNA gel recovery kit (Zymo Research, Irvine, CA, USA). Purified bands were ligated into pGEMT vector using pGEMT easy vector system (Promega, Madison, WI, USA) and transformed in *E. coli* DH5α. Obtained colony were screened through PCR using M13 forward and reverse primers with OneTaq DNA polymerase and sent to sanger sequencing to Biofab s.r.l (Rome, Italy). A list of primers used for the PCR reactions is shown in Suppl. Table S3**.** Plasmids were digested with *Eco*ICRI restriction enzyme (Thermo-Fisher scientific, Waltham, MA, USA) as suggested by manufacturers. Radio labeled RNA probes were synthesized in vitro using T7 polymerase from maxiscript kit (Invitrogen, Carlsbad, CA, USA) following manufacturer instructions and using P^{32} labeled Uridine and as template the EcoICRI-linearized plasmid; after the transcription DNA template was removed through TurboDNAse (Invitrogen, Carlsbad, CA, USA) treatment for 20 minutes at room temperature and probes were cleaned on column using RNA gel recovery kit (Zymo Research, Irvine, CA, USA).

3.5.7 Rapid Amplification of cDNA Ends (RACE)

5’ and 3’ RACE were performed to complete the sequence of Cryphonectria parasitica sclerotimonavirus 1 following a methodology previously described. From total RNA extracted as described, PCR reactions were performed using specific primers amplifying CpSV1 ends: for 5’ RACE a reverse primer annealing specifically 300 bp from the putative end was used for cDNA synthesis and for 3’ RACE a forward primer annealing 300 bp from the putative end was used for cDNA synthesis (Suppl. Table S3). Terminal Deoxynucleotidyl Transferase (Promega, Madison, WI, USA) was used as suggested by the manufacturer to add dATP and dGTP to the obtained tailed cDNA, half of the reaction was added with dATP and half with dGTP. A PCR reaction was then performed using a polyT primer for dATP reactions and polyC primer for dGTP reaction together with a primer specific for CpSV1 and amplifying around 250 bp from the putative ends (Suppl. Table S3). PCR bands obtained were purified and cloned as already explained. To reconstruct the full sequence of CpSV1 we sequenced clones from dATP and dGTP reactions to determine unambiguously which is the last nucleotide of the viral genome on each end.

3.5.8 Northern blot analysis

To perform Northern blot analysis on CpaV1 RNA, total RNA from infected and non-infected fungi was separated on Hepes-EDTA agarose gel after denaturation with deionized glyoxal and DMSO as described in literature (Ferriol *et al.*, 2018). RNA gels

were blotted on Immobilon-Ny+ membrane (Merck, Darmstadt, Germany) and blots were hybridized with radioactively labelled probes at 60° C using ULTRAhyb™ Ultrasensitive Hybridization Buffer (Thermo-Fisher scientific, Waltham, MA, USA) following manufacturer instruction.

3.5.9 RNAse R treatment

To investigate if CpaV1 genome is circular RNA we treated extracted RNAs with RNAse R (Lucigen, Middleton, WI, USA). We setup the reactions putting 15 µg of ACP 34 RNA in a 20 µl reaction containing 2 µl of 10X reaction buffer for RNAse R and collecting 5 µl of the mix as a control before adding the enzyme. 1 µl of RNAse R was added to the remaining mix and solution was incubated at 37° C and 5 µl were collected after 30 minutes, 1 hour and 2 hours. After each collection, RNAse R was inactivated through 5 minutes incubation at 80 °C. All the different timepoints of the reactions were analyzed by northern blot as described above.

3.5.10 Single conidia colonies

Conidia from two month old plates grown at 26 °C and with a light/dark cycle of 12 hours were harvested with 15 ml of sterile water on the petri and scrubbing the mycelia. Liquid was then collected and filtered to obtain conidia. Conidia were then plated in serial dilutions and single conidia colonies were transferred after 3 days. CpaV1 presence was then checked through q-RT-PCR as described above.

3.5.11 Phylogenetic analysis

Phylogenetic analysis was performed on Cryphonectria parasitica sclerotimonavirus 1 RdRP sequence to describe phylogenetic relations of the virus with those already characterized and available in the databases. ORF prediction on the viral contig was performed with ORFinder and predicted RdRP sequence was blasted to retrieve the closest viral sequence in the database. After choosing the RdRPs to include in the analysis, sequence alignment was performed using MUSCLE implemented in MEGA 6.0 (Edgar, 2004). Obtained alignment was exported in fasta format and submitted to IqTREE webserver (Trifinopoulos *et al.*, 2016) to build the phylogenetic tree using ML method and performing automatic model selection. Ultrafast bootstrap analysis was performed by IqTREE with 1000 replicates. Branches under a bootstrap value of 50 were collapsed to polytomy. Sequences used for the phylogenetic analysis are shown in Suppl. Table S5.

3.5.12 Sequence availability

Sequence analyzed in this study can be found on NCBI database with this accession numbers: Cryphonectria parasitica sclerotimonavirus 1 (MT354565), Cryphonectria parasitica ambivirus 1 (MT354566), Rhizoctonia solani ambivirus 1 and 2 (MT354567, MT354568). Illumina reads generated in the study are submitted on NCBI with the BioProject number PRJNA625953, biosamples: SAMN14612183, SAMN14612184.

3.6 Acknowledgement

This work was partially supported by the program within the framework of the Joint Bilateral Agreement CNR/Azerbaijan National Academy of Sciences (N.0003795, 22/01/2016). M.F. is supported by a PhD fellowship from the University of Turin. Grateful thanks to Dr. Stefania Daghino for critically reading and discussing the manuscript.

3.7 Author contributions

MF performed all the bioinformatic analysis and the molecular characterization of the identified viruses. EI isolated the fungal strains and performed the RNA extraction. HI contributed to the initial collection. AD and MT supervised the work, provided funding and edited the manuscript.

3.8 Figures and tables

Figure 1. **Location of *Cryphonectria parasitica* sampling districts in Azerbaijan.** In 2018 and 2019, three locations, Zagatala (1), Shaki (2), Qabala (3) were sampled for *C. parasitica* virome characterization.

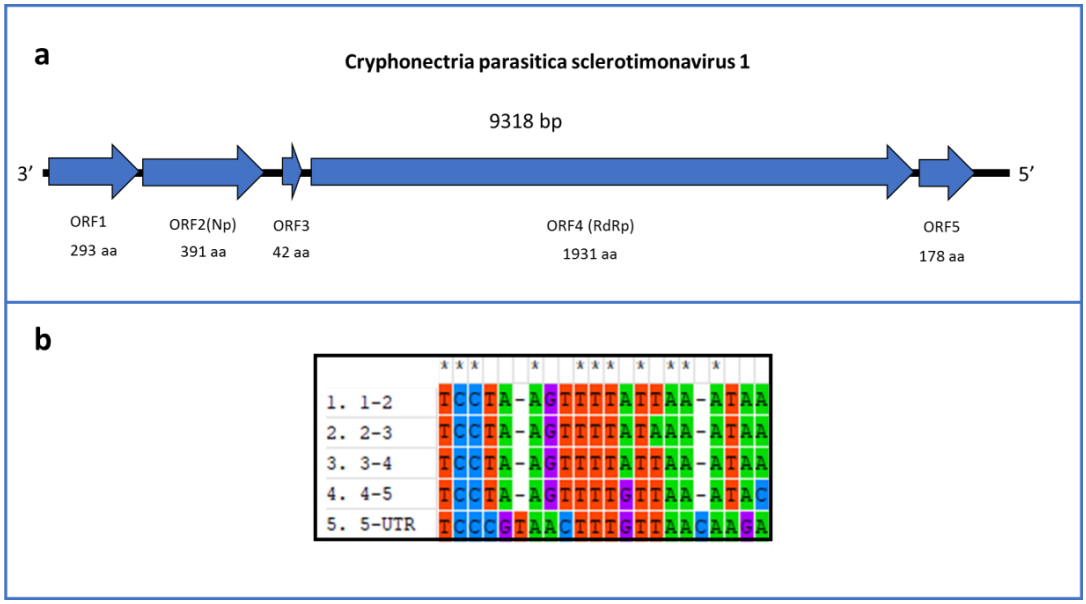


Figure 2. Schematic representation of genome organization of Cryphonectria parasitica sclerotimonavirus 1 and alignment of inter-ORFs sequences. Panel **a**: CpSV1 genome organization. Blue arrows represent open reading frames (ORF)s. A continuous black line represent the genome. Panel **b**: sequence alignment showing conserved RNA regions between ORFs. The left column indicates the ORFs flanking the sequence.

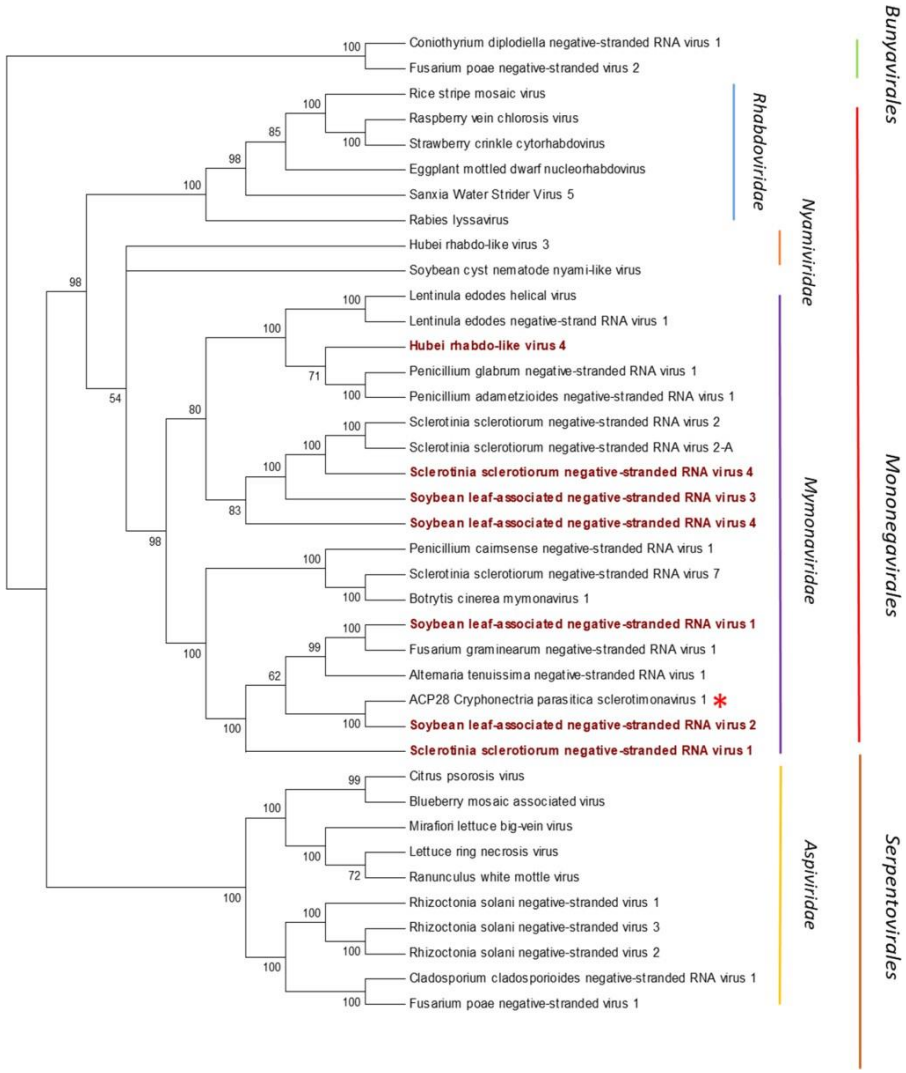


Figure 3. Phylogenetic analysis of negative strand RNA viruses closely related to Cryphonectria parasitica sclerotimonavirus 1 (CpSV1). Orders and families are underlined with different colors. Viruses belonging to Mymonaviruses species accepted by ICTV are shown in red. CpSV1 is marked by a red asterisk. Branches with bootstrap value lower than 50 were collapsed. The specific methodology used to build the Maximum Likelyhood phylogenetic tree is explained in the method section and accession numbers of viruses contained in the analysis are found in Supplementary table 5.

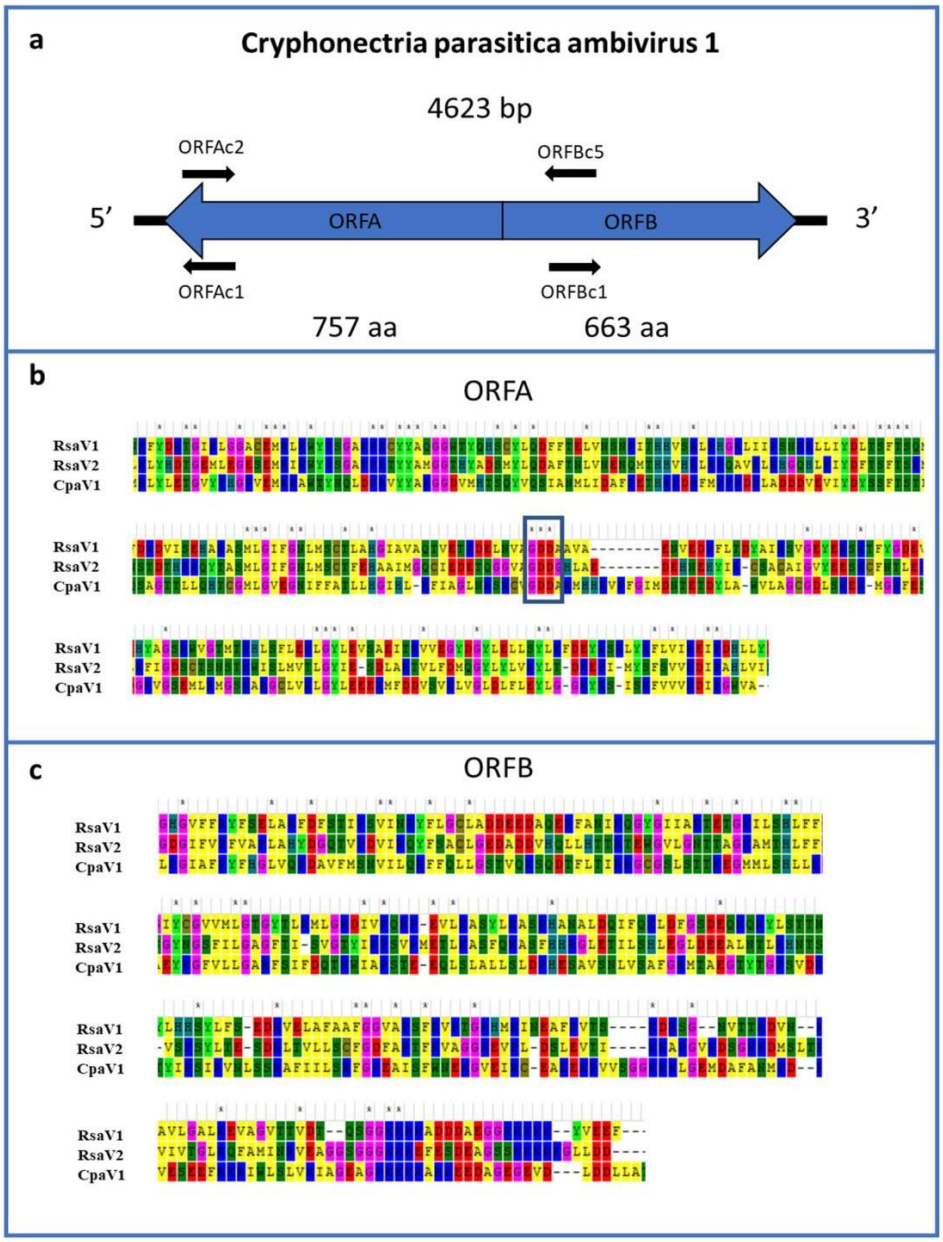

a
Cryphonectria parasitica ambivirus 1
4623 bp
ORFAc2
ORFBc5
5'
ORFA
ORFB
3'
ORFAc1
ORFBc1
757 aa
663 aa
b
ORFA
RsaV1
RsaV2
CpaV1
c
ORFB
RsaV1
RsaV2
CpaV1

Figure 4. Genome organization of Cryphonectria parasitica ambivirus 1 (CpaV1) and amino acid conservation of the putative homologous proteins. Panel **a** shows a schematic representation of Cryphonectria parasitica ambivirus 1, ORFs and their orientation on the sequence are represented by blue arrows**.** Black arrows show orientation and position of the northern blot probes. A black continuous line represents the genome. Panel **b** show parts of an alignment between CpaV1 ORFA (containing the GDD triad) and the conserved proteins retrieved from Rhizoctonia solani ambivirus 1 and 2. Panel **c** displays parts of an alignment between CpaV1 ORFB and the conserved proteins retrieved from Rhizoctonia solani ambivirus 1 and 2. In panels **b** and **c** conserved amino acid residues are marked by the asterisks.

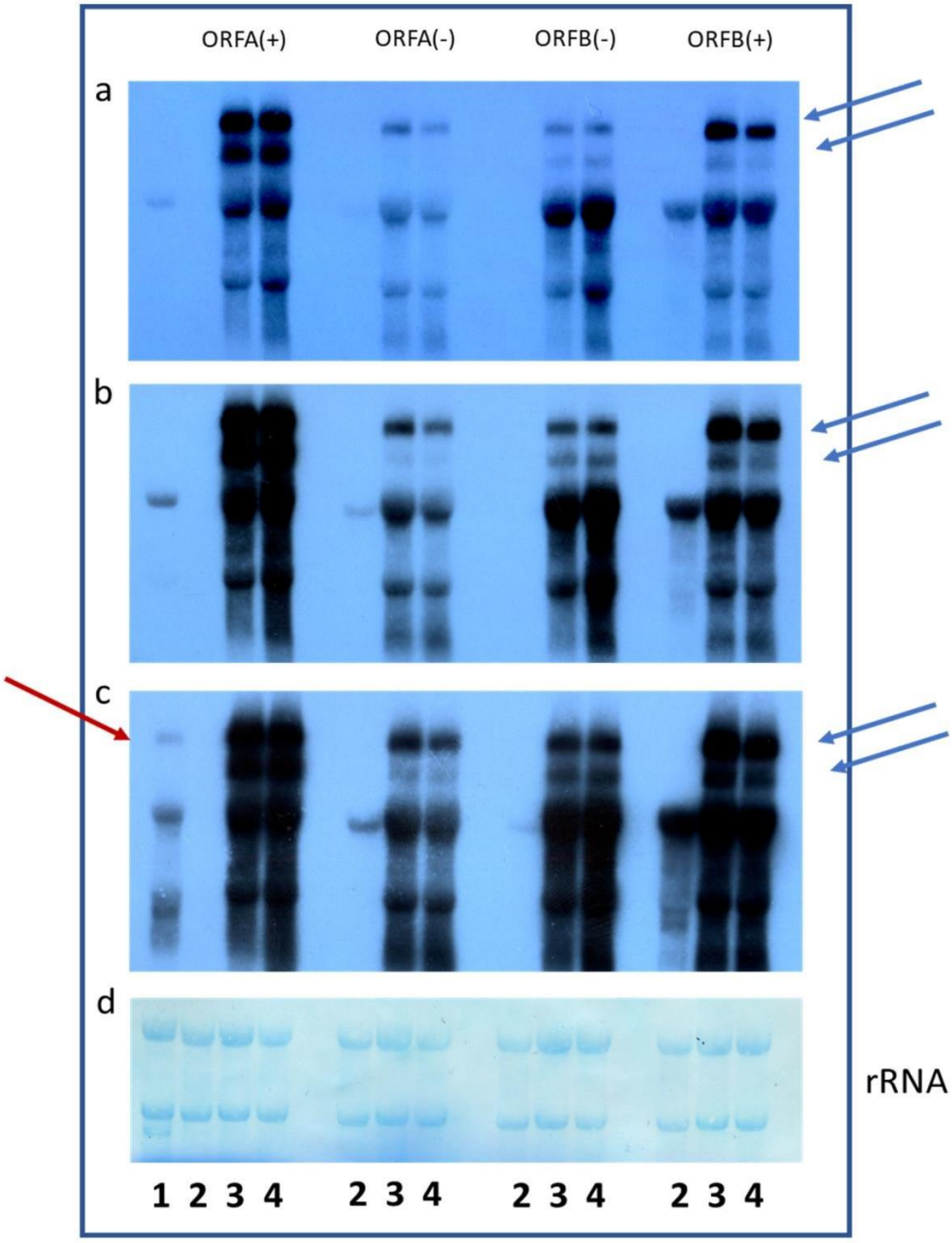

ORFA(+)
ORFA(-)
ORFB(-)
ORFB(+)
a
b
c
d
rRNA
1 2 3 4
2 3 4
2 3 4
2 3 4

Figure 5. Northern blot analysis of Cryphonectria parasitica ambivirus 1 (CpaV1).
Samples are indicated by numbers at the bottom of the panel: 1= Total RNA from tomato infected with tomato brown rugose fruit virus (TBRFV). 2= RNA from fungal isolate ACP 23. 3= RNA from ACP 34. 4= RNA from isolate ACP 36. Panel **a** show result after an exposure of 2 hours. Panel **b** shows results after an exposure of 5 hours. Panel **c** shows results after adding a probe specific for Tomato brown rugose fruit virus and exposure the blot for 5 hours. Panel **d** shows ribosomal RNA (rRNA) staining with methylene blue of the blot before hybridization. Probes used on each sample repetition are shown in the upper part of the picture. 3 µg of RNA was used for each sample. Blue arrows point to the specific doublet of CpAV1. Red arrows points to the 6.3 kB RNA of TBRFV.

Table 1. Viruses found in Azerbaijani *C. parasitica* populations. Information about NCBI accession number, genome length and reads mapping on the viral genome are shown.

Virus	GeneBankID	Genome lenght (bp)	Number of Reads
Cryphonectria parasitica sclerotimonavirus 1	MT354565	9318	32938
Cryphonectria parasitica ambivirus 1	MT354566	4623	542088 (NGS Sample1) 217680 (NGS Sample 2)

Table 2. Information about first blastp hit for each predicted protein encoded by Cryphonectria parasitica sclerotimonavirus 1 and Cryphonectria parasitica ambivirus 1. Blastp search was conducted using PSI-BLAST algorithm.

Virus		Blastp First Hit	Query cover	E value	Per. Ident	Accession
Cryphonectria parasitica sclerotimonavirus 1	ORF1	gp1 [Sclerotinia sclerotiorum negative-stranded RNA virus 3]	38%	6.00E-09	31.75%	YP_009129263.1
	ORF2	ORF2 [Plasmopara viticola associated mononega virus 2]	99%	1.00E-149	56.15%	QHD64786.1
	ORF3	No significant similarity found	/	/	/	/
	ORF4	RdRp [Soybean leaf-associated negative-stranded RNA virus 2]	99%	0.0	61.85%	ALM62227.1
	ORF5	ORF4 [Plasmopara viticola associated mononega virus 2]	89%	1.00E-11	33.33%	QHD64788.1
Cryphonectria parasitica ambivirus 1	ORFA	hypothetical protein [Agaricus bisporus]	44%	8.00E-13	24.93%	AQM32755.1
	ORFB	hypothetical protein GcM3_200062 [Golovinomyces cichoracearum]	93%	8.00E-128	39.94%	RKF55685.1

3.9 References

Aghayeva D., Harrington T. (2007) First report of Cryphonectria parasitica on chestnut (Castanea sativa) in Azerbaijan. *New Dis. Rep,* **15**, 33.

Aghayeva D. N., Rigling D., Prospero S. (2017) Low genetic diversity but frequent sexual reproduction of the chestnut blight fungus Cryphonectria parasitica in Azerbaijan. *Forest Pathology,* **47**.

Ayllon M. A., Turina M., Xie J., Nerva L., Marzano S.-Y. L., Donaire L., Jiang D., Consortium I. R. (2020) ICTV Virus Taxonomy Profile: Botourmiaviridae. *Journal of General Virology,* **101**, 454-455.

Bertran A., Ciuffo M., Margaria P., Rosa C., Resende R. O., Turina M. (2016) Host-specific accumulation and temperature effects on the generation of dimeric viral RNA species derived from the S-RNA of members of the Tospovirus genus. *Journal of General Virology,* **97**, 3051-3062.

Bian R. L., Andika I. B., Pang T. X., Lian Z. Q., Wei S., Niu E. B., Wu Y. F., Kondo H., Liu X. L., Sun L. Y. (2020) Facilitative and synergistic interactions between fungal and plant viruses. *Proceedings of the National Academy of Sciences of the United States of America,* **117**, 3779-3788.

Buchfink B., Xie C., Huson D. H. (2015) Fast and sensitive protein alignment using DIAMOND. *Nature Methods,* **12**, 59-60.

Chiapello M., Rodríguez-Romero J., Ayllón M. A., Turina M. (2020) Analysis of the virome associated to grapevine downy mildew lesions reveals new mycovirus lineages. *Virus Evolution*, veaa058, https://doi.org/10.1093/ve/veaa058

Crouch J. A., Dawe A., Aerts A., Barry K., Churchill A. C. L., Grimwood J., Hillman B., Milgroom M. G., Pangilinan J., Smith M., Salamov A., Schmutz J., Yadav J., Grigoriev I. V., Nuss D. (2020) Genome sequence of the chestnut blight fungus Cryphonectria parasitica EP155: A fundamental resource for an archetypical invasive plant pathogen. *Phytopathology*.

Dawe A. L., Nuss D. L. (2013) Hypovirus Molecular Biology: From Koch's Postulates to Host Self-Recognition Genes that Restrict Virus Transmission. *Advances in Virus Research, Vol 86: Mycoviruses,* **86**, 109-147.

Deakin G., Dobbs E., Bennett J. M., Jones I. M., Grogan H. M., Burton K. S. (2017) Multiple viral infections in Agaricus bisporus - Characterisation of 18 unique RNA viruses and 8 ORFans identified by deep sequencing. *Scientific Reports,* **7**.

Donaire L., Ayllon M. A. (2017) Deep sequencing of mycovirus-derived small RNAs from Botrytis species. *Molecular Plant Pathology,* **18**, 1127-1137.

Edgar R. C. (2004) MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Research,* **32**, 1792-1797.

Eusebio-Cope A., Sun L. Y., Tanaka T., Chiba S., Kasahara S., Suzuki N. (2015) The chestnut blight fungus for studies on virus/host and virus/virus interactions: From a natural to a model host. *Virology,* **477**, 164-175.

Ferriol I., Vallino M., Ciuffo M., Nigg J. C., Zamora-Macorra E. J., Falk B. W., Turina M. (2018) The Torradovirus-specific RNA2-ORF1 protein is necessary for plant systemic infection. *Molecular Plant Pathology,* **19**, 1319-1331.

Gorbalenya A. E., Pringle F. M., Zeddam J. L., Luke B. T., Cameron C. E., Kalmakoff J., Hanzlik T. N., Gordon K. H. J., Ward V. K. (2002) The palm subdomain-based active site is internally permuted in viral RNA-dependent RNA polymerases of an ancient lineage. *Journal of Molecular Biology,* **324**, 47-62.

Grente J. (1965) Les formes hypovirulentes d'Endothia parasitica et les espoirs de lutte contre le chancre du chataignier. *CR Acad. Agric. France,* **51**, 1033-1037.

Haas B. J., Papanicolaou A., Yassour M., Grabherr M., Blood P. D., Bowden J., Couger M. B., Eccles D., Li B., Lieber M. (2013) De novo transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. *Nature protocols,* **8**, 1494-1512.

Hillman B. I., Cai G. (2013) The Family Narnaviridae: Simplest of RNA Viruses. *Advances in Virus Research, Vol 86: Mycoviruses,* **86**, 149-176.

Hillman B. I., Suzuki N. (2004) Viruses of the chestnut blight fungus, Cryphonectria parasitica. *Advances in Virus Research, Vol. 63,* **63**, 423-472.

Langmead B., Salzberg S. L. (2012) Fast gapped-read alignment with Bowtie 2. *Nature Methods,* **9**, 357-U354.

Li H., Handsaker B., Wysoker A., Fennell T., Ruan J., Homer N., Marth G., Abecasis G., Durbin R., Genome Project Data P. (2009) The Sequence Alignment/Map format and SAMTOOLS. *Bioinformatics,* **25**, 2078-2079.

Liu L., Xie J., Cheng J., Fu Y., Li G., Yi X., Jiang D. (2014) Fungal negative-stranded RNA virus that is related to bornaviruses and nyaviruses. *Proceedings of the National Academy of Sciences of the United States of America,* **111**, 12205-12210.

Marzano S.-Y. L., Nelson B. D., Ajayi-Oyetunde O., Bradley C. A., Hughes T. J., Hartman G. L., Eastburn D. M., Domier L. L. (2016) Identification of Diverse Mycoviruses through Metatranscriptomics Characterization of the Viromes of Five Major Fungal Plant Pathogens. *Journal of Virology,* **90**, 6846-6863.

Milgroom M. G., Cortesi P. (2004) Biological control of chestnut blight with hypovirulence: A critical analysis. *Annual Review of Phytopathology,* **42**, 311-338.

Milgroom M. G., Lipari S. E. (1995) Spatial-analysis of nuclear and mitochondrial RFPL genotypes in populations of the chestnut blight fungus, *Cryphonectria parasitica*. *Molecular Ecology,* **4**, 633-642.

Milne I., Bayer M., Stephen G., Cardle L., Marshall D. (2016) Tablet: visualizing next-generation sequence assemblies and mappings. *In: Plant Bioinformatics*, pp. 253-268. Springer.

Nerva L., Ciuffo M., Vallino M., Margaria P., Varese G. C., Gnavi G., Turina M. (2016) Multiple approaches for the detection and characterization of viral and plasmid symbionts from a collection of marine fungi. *Virus Research,* **219**, 22-38.

Nerva L., Forgia M., Ciuffo M., Chitarra W., Chiapello M., Vallino M., Varese G. C., Turina M. (2019) The mycovirome of a fungal collection from the sea cucumber Holothuria polii. *Virus research,* **273**, 197737-197737.

Nerva L., Silvestri A., Ciuffo M., Palmano S., Varese G. C., Turina M. (2017) Transmission of Penicillium aurantiogriseum partiti-like virus 1 to a new fungal host (Cryphonectria parasitica) confers higher resistance to salinity and reveals adaptive genomic changes. *Environmental Microbiology,* **19**, 4480-4492.

Nerva L., Varese G. C., Turina M. (2018) Different Approaches to Discover Mycovirus Associated to Marine Organisms. *Methods in molecular biology (Clifton, N.J.),* **1746**, 97-114.

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

Picarelli M., 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**.

Prospero S., Lutz A., Tavadze B., Supatashvili A., Rigling D. (2013) Discovery of a new gene pool and a high genetic diversity of the chestnut blight fungus Cryphonectria parasitica in Caucasian Georgia. *Infection Genetics and Evolution,* **20**, 131-139.

Puhalla J. E., Anagnostakis S. L. (1971) Genetic and nutritional requirements of Endothia parasitica. *Phytopathology,* **61**, 169-173.

Rigling D., Borst N., Cornejo C., Supatashvili A., Prospero S. (2018) Genetic and Phenotypic Characterization of Cryphonectria hypovirus 1 from Eurasian Georgia. *Viruses-Basel,* **10**.

Rigling D., Prospero S. (2017) Cryphonectria parasitica, the causal agent of chestnut blight: Invasion history, population biology and disease control. *Molecular Plant Pathology*.

Rigling D., Prospero S. (2018) Cryphonectria parasitica, the causal agent of chestnut blight: invasion history, population biology and disease control. *Molecular Plant Pathology,* **19**, 7-20.

Salaipeth L., Chiba S., Eusebio-Cope A., Kanematsu S., Suzuki N. (2014) Biological properties and expression strategy of rosellinia necatrix megabirnavirus 1 analysed in an experimental host, Cryphonectria parasitica. *Journal of General Virology,* **95**, 740-750.

Salem N., Mansour A., Ciuffo M., Falk B. W., Turina M. (2016) A new tobamovirus infecting tomato crops in Jordan. *Archives of virology,* **161**, 503-506.

Sato Y., Shamsi W., Jamal A., Bhatti M. F., Kondo H., Suzuki N. (2020) Hadaka Virus 1: a Capsidless Eleven-Segmented Positive-Sense Single-Stranded RNA Virus from a Phytopathogenic Fungus, Fusarium oxysporum. *mBio,* **11**.

Suzuki N., Ghabrial S. A., Kim K. H., Pearson M., Marzano S. Y. L., Yaegashi H., Xie J. T., Guo L. H., Kondo H., Koloniuk I., Hillman B. I., Consortium I. R. (2018) ICTV Virus Taxonomy Profile: Hypoviridae. *Journal of General Virology,* **99**, 615-616.

te Velthuis A. J. W. (2014) Common and unique features of viral RNA-dependent polymerases. *Cellular and Molecular Life Sciences,* **71**, 4403-4420.

Trifinopoulos J., Lam-Tung N., von Haeseler A., Minh B. Q. (2016) W-IQ-TREE: a fast online phylogenetic tool for maximum likelihood analysis. *Nucleic Acids Research,* **44**, W232-W235.

Turina M., Rostagno L. (2007) Virus-induced hypovirulence in Cryphonectria parasitica: still an unresolved conundrum. *Journal of Plant Pathology,* **89**, 165-178.

Untergasser A., Cutcutache I., Koressaar T., Ye J., Faircloth B. C., Remm M., Rozen S. G. (2012) Primer3-new capabilities and interfaces. *Nucleic Acids Research,* **40**.

Wibberg D., Jelonek L., Rupp O., Hennig M., Eikmeyer F., Goesmann A., Hartmann A., Borriss R., Grosch R., Puhler A., Schluter A. (2013) Establishment and interpretation of the genome sequence of the phytopathogenic fungus Rhizoctonia solani AG1-IB isolate 7/3/14. Journal of Biotechnology, **167**, 142-155.

Wolf, Y.I., Kazlauskas, D., Iranzo, J., Lucia-Sanz, A., Kuhn, J.H., Krupovic, M., Dolja, V.V., Koonin, E.V., 2018. Origins and Evolution of the Global RNA Virome. Mbio **9**.

Chapter 4

The virome from a collection of endomycorrhizal fungi reveals new viral taxa with unprecedented genome organization

Suvi Sutela[1,†], Marco Forgia[2,†], Eeva J. Vainio[1], Marco Chiapello[2], Stefania Daghino[3], Marta Vallino[2], Elena Martino[3], Mariangela Girlanda[3], Silvia Perotto[3] and Massimo Turina[2]

†First and second author equally contributed to the work

1 Natural Resources Institute Finland (Luke), Forest Health and Biodiversity Group, Latokartanonkaari 9, Helsinki FI-00790, Finland
2 Institute for Sustainable Plant Protection, CNR, Strada delle Cacce 73, Torino 10135, Italy
3 Department of Life Science and Systems Biology, University of Torino, Viale Mattioli 25, Torino 10125, Italy

Published on Virus Evolution
DOI: 10.1093/ve/veaa076

4.1 Abstract

Mutualistic plant-associated fungi are recognized as important drivers in plant evolution, diversity and health. The discovery that mycoviruses can take part and play important roles in symbiotic tripartite interactions has prompted us to study the viromes associated with a collection of ericoid and orchid mycorrhizal (ERM and ORM, respectively) fungi. Our study, based on high-throughput sequencing of transcriptomes (RNAseq) from fungal isolates grown in axenic cultures, revealed in both ERM and ORM fungi the presence of new mycoviruses closely related to already classified virus taxa, but also new viruses that expand the boundaries of characterized RNA virus diversity to previously undescribed evolutionary trajectories. In ERM fungi, we provide first evidence of a bipartite virus, distantly related to narnaviruses, that splits the RNA-dependent RNA polymerase (RdRP) palm domain into two distinct proteins, encoded by each of the two segments. Furthermore, in one isolate of the ORM fungus *Tulasnella* spp. we detected a 12 kb genomic fragment coding for an RdRP with features of bunyavirus-like RdRPs. However, this 12 kb genomic RNA has the unique features, for *Bunyavirales* members, of being tri-cistronic and carrying ORFs for the putative RdRP and putative nucleocapsid in ambisense orientation on the same genomic RNA. Finally, a number of ORM fungal isolates harbored a group of ambisense bicistronic viruses with a genomic size of around 5 kb, where we could identify a putative RdRP palm domain that has some features of plus strand RNA viruses; these new viruses may represent a new lineage in the Riboviria, as they could not be reliably assigned to any of the branches in the recently derived monophyletic tree that includes most viruses with an RNA genome.

Keywords
Viruses, ericoid mycorrhizal fungi, orchid mycorrhizal fungi, RNA-dependent RNA polymerase

4.2 Introduction

Viruses affect our lives in pervasive ways, as proved by the current SARS-CoV-2-caused pandemic (Wu et al. 2020) and, up to a decade ago, knowledge of virus biodiversity—the virosphere—was strongly biased by focus on viruses that affect our health, or that cause economic damage by infecting our plant crops, or the animals we have domesticated. Such anthropocentric view of viruses as the ultimate pathogens neglects their basic influence in the evolution of life as we know it (Ryan 2009), or their immense effects on ecological systems as controller of microbial populations (Rohwer et al. 2009), particularly in the oceans (Suttle 2007). Describing viral diversity and the evolutionary relationships among viruses is not a mere classificatory exercise, but it has profound impacts, for example to elucidate their origins and the origin of life (Krupovic et al. 2019; Krupovic et al. 2020). In practice, virome studies may be used to identify host reservoirs involved in viral spillovers to new species, mostly among vertebrates (Mollentze & Streicker 2020), or to help assess the risk of using specific viruses as biological tools with respect to their host specificity or off-target effects. Our knowledge of the virosphere changed enormously with the contribution of high throughput sequencing (HTS) techniques to characterize viromes associated with different hosts and metagenomic samples in an unbiased approach, independently of their pathogenic potential.

All the viruses with an RNA genome so far characterized share a viral polymerase protein carrying a common domain, called palm domain, at the core of the polymerase catalytic activity (te Velthuis 2014). This domain is shared by both reverse transcriptases and the RNA-dependent RNA polymerases (RdRP); the latter are proteins that amplify viral genomes synthesizing plus and minus stranded RNA from an RNA template. There is now a consensus that known viruses with RNA genomes are monophyletic, and this resulted in a recent effort to reconstruct their evolutionary trajectory in a single phylogenetic tree (Wolf et al. 2018) recognized by the International Committee on Taxonomy of Viruses (ICTV), which currently includes all viruses with an RNA genome in the realm *Riboviria*, and in particular those that use an RdRP for their replication are placed in the kingdom *Orthornavirae* (Koonin et al. 2020). The greater contribution to the characterization of the diversity of the RNA virosphere comes from studies on the viromes of invertebrates, mainly from arthropods (Li et al. 2015; Shi et al. 2016; Kaefer et al. 2019). The same HTS techniques have been applied to the characterization of viromes associated with fungi, but the number of fungal hosts investigated is still very low (Vainio et al. 2015; Marzano et al. 2016; Donaire & Ayllon 2017; Nerva et al. 2019a) and their contribution to the overall diversity of fungal viruses (mycoviruses) is limited. A more traditional approach based on the characterization of individual viruses has nevertheless allowed the identification of new classes of mycoviruses, such as the first negative strand mycovirus (Liu et al. 2014), the first ssDNA mycovirus (Yu et al. 2010) or new polysegmented RNA mycoviruses with limited resemblance to existing RNA viruses (Kanhayuwa et al. 2015; Sato et al. 2020).

Mycoviruses were discovered in the early 1960's as a cause of disease in the cultivated mushroom *Agaricus bisporus* (Hollings 1962) and were later found in most fungal taxa investigated (Ghabrial et al. 2015). Although fungi occupy virtually all ecological niches, mycoviruses of plant-interacting fungi are particularly significant for crop protection because they can influence the phenotype of their host. In particular, mycoviruses that cause hypovirulence in phytopathogenic fungi have attracted a lot of interest over the last few decades for their possible use as biocontrol agents (Pearson et al. 2009). In addition, fungi that interact with plants are of particular interest as recent studies have shown that, in virus-fungus-plant tripartite interactions, mycoviruses could spread beyond their native fungal hosts and be transferred to host plants by horizontal virus transfers (HVTs); similarly, plant viruses can be transferred to plant-interacting fungi. Indirect evidence of such HVTs derives from phylogenetic analysis of some RNA virus clades, which include closely related plant viruses and mycoviruses, as is the case for the *Partitiviridae*, the *Endornaviridae* and the *Mitoviridae*, suggesting a relatively recent cross-kingdom virus transmission (Roossinck 2019). Such HVTs often resulted in cryptic persistent viral plant infections that might provide specific beneficial traits (Takahashi et al. 2019). Evidence of cross-kingdom HVT was indeed shown for plant viruses replicating in fungi both experimentally (Mascia et al. 2019) and naturally (Andika et al. 2017), and for mycoviruses replicating in plant cells (Nerva et al. 2017) and in whole plants (Bian et al. 2020).

So far, most studies on mycoviruses of plant-interacting fungi have focused on phytopathogenic fungi as model systems, but other biological systems could have an even greater significance to investigate the interplay among mycoviruses, fungi and plant hosts. In addition to fungal pathogens, plants interact in fact with a broad spectrum of fungal endophytes that reside within the plant tissues as symbionts, often contributing to plant growth and/or defense against biotic and abiotic stress (Gange et al. 2019). Based on their life histories, taxonomic position and colonized host plants/organs, fungal endophytes have been classified into four groups (Rodriguez et al. 2009), and mycoviruses have been found in all four groups (Herrero et al. 2009; Bao & Roossinck 2013). Another taxonomically heterogenous group of fungi in close relationship with plants are the mycorrhizal fungi, which differ from endophytes because they form recognizable fungal inter- or intracellular structures within the root tissues. In ectomycorrhizas, mainly formed by basidiomycetes and some ascomycetes, the fungal partner remains confined in the intercellular spaces of the host root tissues. In endomycorrhizas, by contrast, the symbiotic fungal hyphae colonize the root cells and form specialized intracellular plant-fungus interfaces that increase bidirectional exchanges between partners (Bonfante and Genre 2010). Fungi in the subphylum Glomeromycotina form highly branched fungal structures called arbuscules in the majority of angiosperms, giving rise to the arbuscular mycorrhiza, whereas distinct lineages of asco- and basidiomycetes form coiled hyphal structures in specific families of angiosperms, namely the Ericaceae and the Orchidaceae, giving rise to the ericoid mycorrhiza (ERM) and the orchid mycorrhiza (ORM), respectively. ERM fungi are mainly ascomycetes belonging to the class Leotiomycetes, but include some basidiomycetes in the Serendipitaceae family as well (Weiss et al. 2016). ERM fungi interact with the host

plant but can also grow as saprotrophs in the soil, where they can degrade a wide range of organic substrates (Smith & Read 2008), thus contributing to the mobilization of nutrients (Read & Stribley 1973) and to the ecological success of their host plants (Smith & Read 2008).

ORM fungi are essential for orchid propagation in nature, since orchids produce minute seeds lacking stored energy sources, which are unable to germinate and develop further unless they are colonized by symbiotic fungi that provide the host with nutrients, including organic carbon. Orchids are usually found to associate with fungi of three families in the Basidiomycota, namely the Tulasnellaceae and Ceratobasidiaceae (Cantharellales) and the Serendipitaceae (Sebacinales) (Dearnaley et al. 2012). These families were previously assigned to the form-genus *Rhizoctonia*.

Mycorrhizal fungi, which contribute profoundly to the growth and health of plant partners, have established a long co-evolution with the roots of most terrestrial plants (Smith & Read 2008). However, mycoviruses have been reported in less than twenty mycorrhizal fungal genera so far, not even covering all different mycorrhizal types (Blattný & Králík 1968; Dieleman-Van Zaayen et al. 1970; Huttinga et al. 1975; Bai et al. 1997; Stielow and Menzel 2010; Petrzik et al. 2016; Vainio et al. 2017). In particular, no mycoviruses have been reported from ERM fungi so far, and a few mycoviruses have been characterized in ORM fungi from Australia, including a newly described hypovirus and a mitovirus that could not be ascribed to known taxa of the same genera (Ong et al. 2016; Ong et al. 2017; Ong et al. 2018).

To investigate the possible contribution of mycoviruses to the evolution of viruses, we need to increase our knowledge on the diversity and distribution of mycoviruses in mycorrhizal fungi, which have been poorly investigated compared with other fungal groups. The purpose of this work was to characterize the virome of ERM and ORM fungi, involved in two fairly neglected endomycorrhizal types. Endomycorrhizal fungi are particularly interesting fungal hosts to investigate HVT between fungi and plants, and vice-versa, because they maintain close contacts without the onset of structural and molecular host defense, that often can hamper the exchanges in pathogenic plant-fungal interactions. We examined 37 ERM fungal strains, mostly *Oidiodendron maius* isolated from *Vaccinium* species or from *Calluna vulgaris,* and 12 ORM fungal strains in the genera *Tulasnella* and *Ceratobasidium*, isolated from the roots of Mediterranean terrestrial orchids in Italy. The analysis revealed new viral taxa with unprecedented genome organizations, expanding our view of mycovirus diversity, genome structures and phylogenesis of RNA viruses.

4.3 Materials and Methods

4.3.1 Origin of the fungal isolates

The ERM and ORM fungi investigated in this study are listed in Supplementary online Table S1. The 37 ERM fungal strains have been previously isolated from roots of *C. vulgaris*, *V. myrtillus*, *V. corymbosum* or *V. angustifolium*, collected in different countries (Italy, Great Britain, Poland, Canada). Two sampling sites were characterized by either mildly (Vallino et al. 2011) or strongly (Martino et al. 2000; Martino et al. 2003) polluted soils. The 12 ORM fungal strains were isolated in Italy, mainly from two terrestrial meadow orchid species: *Serapias vomeracea* and *Orchis purpurea*. Further characteristics of the sites and the isolation and identification methods have been described previously (see references in Supplementary online Table S1). Morphological identification of the strains was confirmed by the sequencing of the ITS regions, available in public databases. The *R. ericae* (UAMH7375-ICMP18553) and *O. maius* (MUT1381-ATCC MYA-4765) (Martino et al. 2018) genomes are available (Mycocosm Sequencing Project, JGI, USA).

4.3.2 Fungal RNA extraction

For the extraction of total RNA, fungal strains were cultivated in liquid cultures: ERM strains in Czapek mineral medium (pH 5.6) with 2% glucose (Martino et al. 2000) supplemented with MES (3.9 g l^{-1}); ORM fungal strains in 2% malt extract medium. Fungal cultures were kept on an orbital shaker (~100 rpm) at 24 °C in the dark for three to seven days depending on the growth of the strain. Three ERM fungal strains (MUT1371, MUT2998 and MUT3000) were grown on Czapek-glucose 2% plates covered with cellophane membranes at 24 °C. Mycelia was collected by vacuum filtering through Miracloth, dried and subsequently frozen and stored at -80 °C prior freeze drying of 48 h. RNA extraction was conducted from 15–37 mg of freeze-dried fungal mycelia using the Spectrum Plant Total RNA Kit (Sigma-Aldrich). Briefly, lyophilized fungal mycelia were homogenized in 900 µl of lysis solution with ceramic and glass beads in FastPrep24. Homogenization (6.5 m/s 30 sec) was conducted three times and the sample mixture was frozen prior to the final homogenization step. Samples were then centrifuged at maximum speed for 5 min at +4 °C and approximately 700 µl of the solution was transferred to a new tube and re-centrifuged. Supernatant was transferred to a filtration column and RNA isolation was performed following manufacturer's instructions with the following exceptions: the used binding solution volume was 1:1 to the sample volume, the volume of wash solution 1 and 2 was 700 µl and elution of RNA was conducted with 45 µl of elution solution. RNA quantity and quality were determined using NanoDrop 2000 (Thermo Scientific) and the integrity of ribosomal RNA was confirmed by running an aliquot of RNA on an agarose gel.

4.3.3 Library construction and assembly and identification of virus-like sequences

Pooled total RNA samples representing 0.5 µg for each fungal strain were sent to Macrogen Korea precipitated in ethanol. The RIN values of pooled RNAs varied between 8.1 and 8.7 and the ribosomal RNA (rRNA) ratios 1.1-1.8. TruSeq Stranded Total RNA with Ribo-Zero Gold Human/Mouse/Rat (Illumina) was used for removal of rRNA and construction of cDNA library. An Illumina platform was used to generate 101 bp pair-end reads.

Trinity 2.6.6 (Grabherr et al. 2011) was utilized in the de novo assembly and run with R (R Development Core Team 2011) on Taito supercluster (csc.taito.fi). Trimmomatic (Bolger et al. 2014) was run as a Trinity plugin and used in removing adapters and trimming reads. A custom virus protein database was used in the BlastX (NCBI BLAST 2.7.1) homology searches run as parallel BLAST in Taito supercluster. Geneious 10.2.6 (Biomatters Ltd) was used for finding open reading frames (ORFs), aligning contigs as well as mapping the reads against the putative virus contigs (Geneious assembler with medium-low sensitivity). To complement virus discovery using BlastX, we performed an analysis of ORFan-encoding RNA segments (i.e. segments with no detectable homology with fungal or viral proteins, at least 1kb in length, encoding for at least a protein >15kDa, with at least 1000reads/kb mapping the contig, and with abundant accumulation of both positive and negative sense reads).

Motifs and domains were searched with MOTIF search (https://www.genome.jp/tools/motif/). Predicted molecular weight and pI were calculated with ExPASy (https://web.expasy.org/compute_pi/).

Viral contigs of interest were mapped on stranded cleaned libraries to give a more complete overview of the transcriptome complexity. Bowtie2 (Langmead & Salzberg 2012) was used to map reads on viral contigs.

TruSeq Stranded Total RNA technology enables not only whole-transcriptome analysis, but also the precise measurement of strand orientation. SAMtools (Li et al. 2009) extracts the reads mapped on sense or anti-sense viral strands, reporting three numbers for each viral contig: the total reads, the positive sense reads (samtools view -F 0x10) and the negative sense reads (samtools view -f 0x10).

In some analyses, a head-to-tail dimer of the protein was used to map reads and reads mapping across the junction were visualized with Tablet (Milne et al. 2016).

4.3.4 Validation of *in silico* assembly of virus-like sequences and assignment of specific host isolates

One µg of total RNA was used in the generation of cDNA with random hexamer primers and RevertAid M-MuLV reverse transcriptase (Thermo Scientific). Reverse transcription was conducted according to manufacturer's instructions but with initial denaturation of 2 min at 98 °C for RNA and random hexamer primer and subsequently performing transcription using 10 U RiboLock RNase Inhibitor (Thermo Scientific) and 100 U RevertAid M-MuLV RT. Screening of fungal hosts was performed in standard reverse transcriptase PCR (RT-PCR) with putative viral contig specific primer pairs (Supplementary Table S2) and DreamTaq DNA Polymerase (Thermo Scientific). Next, DNA of virus positive strains was extracted (Turina et al. 2003) and used as a template in standard PCR with contig specific primer pairs (Supplementary Table S2) and OneTaq DNA Polymerase (New England BioLabs) in order to determine whether viral amplicon would originate from fungal DNA. For the virus contigs described in this work a qRT-PCR protocol was implemented exactly as previously described (Chiapello et al. 2020) with oligonucleotides displayed in Supplementary online Table S2.

The two most original and previously unreported genome organizations were confirmed through overlapping RT-PCR and quasi full-length RT-PCR using cDNA as a template that was obtained with a mix of random primers and specific primers used with SuperScript IV RT (Thermo Fisher Scientific) at 55 °C following instructions provided by the manufacturer. PCR amplification was performed with Phusion polymerase (NEB), following manufacturer's suggestions, using extension times ranging from 1 to 4 minutes, according to the predicted length of the amplification product. Oligonucleotides used for this purpose are displayed in Supplementary Table S2.

4.3.5 Northern blot analysis

Northern blot analysis was carried out on fungal samples extracted as described above and separated on Glyoxal-Hepes agarose gel system as previously described in detail (Ferriol et al. 2018). Ribo-probes were derived with T7 RNA polymerase transcription from cDNA corresponding to specific regions of the viral genomes using radioactive UTP labeled with P^{32} in the reaction mix. We derived cDNA probes cloning PCR products of circa 200-400 bp in pGEM-T easy vector (PROMEGA). In order to obtain minus sense and plus strand probes, both orientations were screened and confirmed by sequencing. Specific primers and position of the probes are displayed in Supplementary online Table S3 and in each figure legend related to virus organization. Hybridization, washes and exposure to film were carried out using ULTRAhyb® Hybridization Buffer (Ambion) as previously described (Ferriol et al. 2018). When loading RNAs for Northern analysis, a total RNA extract of tomato brown rugose fruit virus (ToBRFV)-infected tomato plants was included as a RNA size marker (genomic RNA of 6.3 kb and sgRNA2 of 0.7 kb). Negative controls were isolates of the same or closely related fungal species (included in the RNAseq libraries), which were tested negative based on the RT-PCRs for the specific virus analyzed by Northern blot.

4.3.6 Phylogenetic analysis

Phylogenetic analysis was carried out by first selecting homologous viral proteins from the databases and identifying a possible outgroup (Supplementary online Table S4). For each distinct set of conserved sequences, alignment was performed with Clustal Omega (Sievers et al. 2011) at the EMBL-EBI services website. Phylogenetic tree were derived with the Maximum likelihood methodology implemented in IQ-tree (Trifinopoulos et al. 2016), choosing the best substitution model (Kalyaanamoorthy et al. 2017) and using the ultrafast bootstrap option (Diep Thi et al. 2018). The substitution model and parameters specific for each tree are detailed in each figure legend.

4.3.7 5' and 3' RACE

In order to determine the 5' and 3' prime terminal sequence of the two narnavirus-like viral segments and the bunyavirales-like viral segment we performed a RACE protocol as previously described in detail (Rastgou et al. 2009). Briefly, cDNA with specific primers (specifically described in Supplementary online Table S2) was synthesized from total RNA extracted as described above, using SuperScript IV Reverse Transcriptase (Therm-Fisher Scientific). RNase H digestion was performed on cDNA and the ssDNA was purified with a DNA purification kit following specific instructions for ssDNA (DNA Clean & Concentration Kit (Zymo Research)). dGTP and dATP were used to add a polyG or polyA tail using rTdT, terminal deoxynucleotidyl transferase (Promega). The final PCR step was carried out with

the Oligo-dC(Eco-bam-dc12) or oligo-dT-V previously described (Rastgou et al. 2009) with specific internal primers (Supplementary online Table S2). We used the same protocol to determine 5' and 3' ends, only changing the virus-specific primers, since the virus we tested accumulated a good amount of both +strand and -strand genomic RNA. PCR product were sequenced directly with specific primers or cloned and sequenced if unspecific amplification products were observed.

4.3.8 RNase R digestion

RNase R digestion was performed on total RNA extracted from lyophilized mycelia as described above. We used as negative control RNA extracted from tomato infected with ToBRFV. Digestion was carried out on 6 mg of total RNA in a reaction of 20 µl with buffer and conditions suggested by the manufacturer (Epicentre). Each digestion contained 10 U of enzymes. Reactions were carried out for 2 hrs in total, interrupted every 30 min to separate 5 µl and after inactivation (65 °C for 20 min) run in glyoxal HEPES gels as described above.

4.3.9 Cell fractionation/virus enrichment protocol.

Twenty grams of fresh mycelia (isolate O4) was harvested from a 3-day liquid culture (500 ml) through filtration with Miracloth (Calbiochem). A row extract (EXT) was obtained adding 200 ml of 0.1 Tris pH 8 buffer with EDTA, Na-sulfite and DIECA and homogenizing it in a beat beater for five 60 second rounds (resting for 2 min between each round). A low speed centrifugation (3,000 RPM in a Sorvall GSA rotor) resulted in a supernatant (S1) and a large pellet that was subsequently resuspended in 10 ml of the same buffer (P1). A second low speed centrifugation (10,000 RPM in a Sorvall GSA rotor, using adaptors for Falcon 50 ml conical tubes) resulted in a supernatant (S2) and a pellet, again resuspended in 2.5 ml of 0.1 Tris pH 8. Finally, the supernatant was centrifuged at 36,000 RPM with a Ti55 Beckman rotor for 150 minutes. This resulted in a supernatant (S3) and a final pellet containing the microsomal fraction (P3). Finally, the pellet was resuspended in 2 ml 0.01 Tris buffer pH 7.0. RNA was extracted from equivalents of the original extracts and qRT-PCR on cDNA originating from each fraction was carried out as described above. The P3 fraction was negatively stained and observed at the electron microscope as previously described (Nerva et al. 2016).

4.4 Results

Our ERM fungal collection included 31 *Oidiodendron maius* strains, isolated either from *Vaccinium* species or from *C. vulgaris* (Supplementary online Table S1) (Read 1974; Couture et al. 1983; Dalpe 1986; Perotto et al. 1996; Lacourt et al. 2000; Martino et al. 2000; Martino et al. 2003; Vallino et al. 2011). The ERM fungi in this study also included two *Rhizoscyphus ericae* strains isolated from *C. vulgaris* roots. This was the first fungal species isolated from ERM roots (Pearson & Read 1973). The other four ericoid mycorrhizal strains were sterile mycelia isolated from C. vulgaris, two of them taxonomically related to the so-called '*Rhizoschyphus ericae* aggregate' and two belonging to the Helotiales.

Our ORM fungal collection included 12 strains in the genera *Tulasnella* and *Ceratobasidium*, all isolated from the roots of Mediterranean terrestrial orchid species in Italy (Girlanda et al. 2011).

A number of ITS rDNA sequences were determined on selected isolates to confirm previous assignments (not shown). After growing the fungal isolates in axenic conditions, HTS on rRNA-depleted total RNA was carried out on five distinct libraries containing eight to eleven pooled samples each (Supplementary online Table S5). Overall, we obtained ca. 103–118 million pair-end reads for each library and the reads are available in the SRA archive through the BioProject accession number PRJNA629308. A bioinformatic pipeline previously described in detail (Chiapello et al. 2020) allowed us to characterize putative viral contigs present in each library *in silico* and trace each contig to the specific sample by RT-PCR (Supplementary online Table S5). Below are the main findings described according to the fungal host.

4.4.1 Mycoviruses in ericoid mycorrhizal fungi

After assembling the reads corresponding to each of the libraries of ERM fungi, a first search of a custom-made viral database revealed three putative viral contigs (Table 1). For each putative viral contig, we first investigated their presence in the library by mapping sequencing reads to the viral contigs. Their number is an indication of relative expression and these viruses were all highly abundant in the libraries with high coverage along their genomes (Table 1). We then assigned each putative viral contig to each host isolate by RT-PCR (see Supplementary online Table S5). In some of the *O. maius* isolates we detected two distinct narnavirus-like sequences and one ourmia-like virus (Fig. 1A).

The ourmia-like sequence (contig DN47822) shares the highest BlastX identity (60.7%; query cover 77%) with Combu positive-strand RNA mycovirus (GenBank accession H990636; unpublished), with an E-value of 0. It encodes a single ORF, predicted to be translated into a protein that has a typical RNA-dependent RNA polymerase (RdRP) domain (PFAM code PF05919) with a predicted molecular weight of 72.1 kDa. Phylogenetic analysis shows that it clusters among members of the classified genus *Scleroulivirus* in the family *Botourmiaviridae* (Fig. 1B). For this virus we suggest the name Oidiodendron maius ourmia-like virus 1 (OmOlV1).

The two narna-like viral contigs, DN37559 and DN43802, both have a typical narnavirus-like genome organization, with the potential to express a single protein of circa 88.2 kDa and 89.6 kDa, respectively. However, closer examination of sequence characteristics and

host ranges provides strong evidence that the two contigs represent the two genome segments of a bisegmented narnavirus, designated here as 'Oidiodendron maius splipalmivirus 1 (OmSPV1)' for reasons described below. In our culture collection, both contigs are present in three *O. maius* isolates originating from different mycorrhizal plants growing in the same non-polluted site and representing different clonal individuals (Supplementary online Table S5). For the first putative narna-like contig (DN37559), the closest hit in the NCBI database is that of the RdRP of Plasmopara viticola lesion associated narnavirus 20 (unpublished, GenBank QIR30299.1) with 48.6% identity and a coverage of 86%. For the second narna-like contig (DN43802) there is a unique hit in the viral database, Beihai narna-like virus 22 (Shi et al. 2016), with 26% identity on a very limited part of the genome (23% coverage) and an E-value of 0.09 (Table 1). Phylogenetic analysis includes DN37559 in a distinct clade with other narna-like unassigned viral sequences (Osaki et al. 2016; Zoll et al. 2018; Nerva et al. 2019a) very distantly related to *Leviviridae* and *Narnaviridae* using both a Maximum Likelihood (Fig. 1B) and a Bayesian methodology (Supplementary online Fig. S1); DN43802 was not included in the dendrogram because the sequence was too divergent from known taxa to allow reliable phylogenetic inference. Domain analysis with MOTIF Search (https://www.genome.jp/tools/motif/) for both DN37559 and DN43802 failed to detect the signature of a typical RdRP. Alignment of DN37559 ORF1-encoded protein with a number of other virus RdRPs (for which the evolutionary divergence value is below an acceptable threshold) confirms the presence of homologues of the motif A and B of the palm domain but the absence of the GDD-carrying C domain at their carboxy terminal (Supplementary online Fig. S2)(te Velthuis 2014).

Given the lack of a complete and recognizable RdRP palm domain, we further characterized these two putative virus segments to exclude that one or more copies of cDNA corresponding to these RNA segments were endogenized in the *O. maius* genome and, at the same time, to confirm that these are RNA viruses and do not have a DNA replication intermediate or a DNA genome. Blast search of the DN37559 and DN43802 sequences in the strain whose genome was previously sequenced (Kohler et al. 2015) failed to retrieve any hit, therefore excluding the possibility of endogenization. In addition, quantitative PCR experiments on total nucleic acid did not amplify a small fragment of the putative viral genome. Thus, we could exclude a DNA intermediate during replication (Fig. 1C). We then performed a RACE protocol to determine the exact ends for both DN37559 and DN43802 genomic RNAs; both contigs shared a poly(U) leader at the 5' end, of different variable length (8-12 nt) according to the different cDNA clones sequenced, with two variable nucleotides as terminal nucleotides. At the 3' end, both viral contigs shared a poly(A) tail of variable length (8-10 nt) with two variable nucleotides as sequence termini.

Noticeably, the region of identical nucleotides in the alignment extends downstream of the 5' poly(U) stretch (28 nt) and upstream of the poly(A) stretch (Fig. 1D) strongly suggesting that the two genomic segments are from the same virus (replicated by the same viral RdRP).

To further characterize the nature of these two segments, we checked the number of positive sense and minus strand reads mapping against the genome (Table 2). Surprisingly, a higher number of reads mapped to the negative strand, suggesting that for these virus-like segments, a higher abundance of negative strand RNA accumulates. As a control, we mapped *O. maius* actin, tubulin and ATPase reads on the putative transcript and found the expected accumulation of—almost exclusively—plus strand reads. Furthermore, we also included in the analysis OmOlV1 and it accumulates more plus strand reads than minus strand, as expected from plus strand genome viruses (Table 2). Northern blot analysis confirmed the abundant accumulation of DN37559 and DN43802, with genomes of the expected size, and a stronger signal of the minus strand genomic RNA compared to the plus strand RNA for both DN37559 and DN43802, more evident for DN43802 (Fig. 2 and Supplementary online Fig. S3).

Finally, given the uniqueness of the DN43802 contig, we searched homologues of this virus segments in some Trinity assembly databases for filamentous fungi from our previous works (Nerva et al. 2016; Nerva et al. 2019a; Nerva et al. 2019b) and from a number of fungal transcriptomes available in public databases. We were able to retrieve four putative DN43802 homologous protein sequences: two from esca disease-associated fungi, one from *Holothuria polii*-associated fungi and one from the basidiomycete *Wallemia sebi* (isolate MUT4935) associated with *Posidonia oceanica* (Nerva et al. 2016; Nerva et al. 2019a; Nerva et al. 2019b). Protein sequence alignment shows that these viral sequences, as well as DN43802, align with the N-terminal region of a much larger putative RdRP of the distantly related Beihai narna-like virus 22 (Shi et al. 2016). This region contains the specific C and D subdomain of the putative RdRP (Fig. 3). Given the possible strict association of the two virus segments in our *O. maius* collection (they are in the same libraries and in the same samples, and with identical 5' and 3' ends), we re-examined the libraries that harbor the homologues of DN43802 (i.e. esca disease-associated fungi, *Holothuria polii*-associated fungi, and *Wallemia sebi* isolate MUT4935) for homologues of DN37559. Indeed, we were able to retrieve the corresponding DN37559-like protein encoding segments (Fig. 3). Finally, we were able to trace, through a specific qRT-PCR assay, the presence of the two virus segments found in the *H. polii*-associated fungi to a specific fungal isolate, *Penicillium stoloniferum* MUT2120.

Overall, in five distinct libraries/samples (Fig. 3) we were therefore able to strictly associate, in a putative single virus genome, two virus segments—one that carries the A and B subdomain of the RdRP palm domain at the carboxy terminal of the protein (DN37559-like) and one that carries the C and D subdomain containing the GDD triplet and a conserved K residue, respectively, at the amino terminal of the protein (DN43802-like). Some other lateral conserved domains were also detected following the nomenclature previously used (subdomain I to VIII, Fig. 3) (Koonin 1991). In conclusion, the results indicate that the DN37559 and DN43802 contigs represented the two genome segments of a single virus named after its unique split palm domain as 'Oidiodendron maius splipalmivirus 1'.

4.4.2 Mycoviruses in orchid mycorrhizal fungi

In the collection of ORM fungal isolates, all belonging to the genera *Ceratobasidium* and *Tulasnella* (Basidiomycota), we found a variety of viruses, some of which are new viruses belonging to putative new species in classified virus families; a detailed description of these viruses is provided in Supplementary on line Appendix-I and related Supplementary online Fig. S4. Briefly, these viruses could be affiliated with virus families *Barnaviridae*, *Endornaviridae* and *Mitoviridae* (Table 1).

Besides these well-known virus taxa, one virus contig (DN37204), only found in a single *Tulasnella* isolate (O7), had a surprising unprecedented genome organization. A 12 kb RNA fragment comprises three ORFs (Fig. 4A): the largest one encodes a putative RdRP of 290,3 kDa calculated molecular mass, and the closest match by BLASTp was with the RdRP of Hubei orthoptera virus 2 (E value 1e-07, identity 23%, query coverage 8%) and Guaroa virus (Orthobunyavirus). Both these viruses are multisegmented, and the RdRP is the only protein encoded by the largest of their genomic segments. This protein carries the typical motif of a bunyavirus RdRP (pfam04196, Bunya RdRP). The second largest protein is encoded by the same genome strand (arbitrarily antisense, since RdRP is antisense in *Bunyavirales*) and has a predicted molecular mass of 129,3 kDa. Although we found no specific hits for this protein in the viral database, it carries the Tom22 motif (a protein of the mitochondrial translocase family), with a relatively high E-value of 0.013. Finally, the third ORF encodes, in the sense orientation, a putative nucleocapsid protein (PF02477, Nucleocapsid Nc protein) of 27.8 kDa, that shares some similarity with the Nc of Largemouth bass bunyavirus (Supplementary online Fig. S5A). For this putative virus, we propose the name Tulasnella bunyavirales-like virus 1 (TuBlV1). Phylogenetic analysis indeed placed this new virus fragment inside the *Bunyavirales*, but its assignment to existing clades is not supported statistically (Fig. 4B).

Given the surprising genome organization for a member of the *Bunyavirales* (three ORFs encoded by a single segment), we searched for evidence that the 12 kb contiguous RNA segment exists as a molecule and was not simply assembled *in silico* due to the 5'–3' complementary sequences present in general in multisegmented members of the *Bunyavirales* (that are sometimes assembled in a single sequence). We therefore amplified overlapping segments through RT-PCR spanning the whole genome length (Fig. 5A) and we could indeed confirm the presence of such predicted contiguous RNA. Furthermore, display of the reads mapping on the full-length genome revealed a rather uniform distribution, with several reads validating the link between ORF1 and ORF2 position, and the intergenic region between ORF2 and ORF3 (Supplementary online Fig. S5B). Northern blot analysis showed indeed the presence of a large genomic segment with each of the three probes, thus confirming the existence of the circa 12 kb RNA genomic segment (Fig. 5B). In one RNA sample, a small RNA specifically hybridizing with the putative minus strand probe was also visible, and its size is consistent with a possible subgenomic RNA (Supplementary online Fig. S6A) corresponding to the nucleocapsid coding RNA. Table 2 shows the number of reads mapping to the genomic (+sense reads) and antigenomic (negative sense reads) for the overall genomic segment and for the three coding regions, separately. Overall, we found more negative sense reads, and only the Nc coding region had an excess of plus strand reads, which suggests a specific

subgenomic RNA as expression strategy for this portion of the genome. Furthermore, a U-rich region of circa 40 nt was found after the putative Nc stop codon, a possible signal for transcription termination.

We also checked, through PCR amplifications on total nucleic acid extracts, for the possibility that this virus could be integrated in the fungal genome, but failed to detect any specific product, thus confirming that this is a putative replicating viral segment depending on viral RdRP activity (Supplementary online Fig. S6B).

Finally, to complement virus discovery using BlastX, we performed an analysis of ORFan-encoding RNA segments (i.e. segments with no detectable homology with fungal or viral proteins). In the XEO library we initially detected two contigs (DN30310 and DN33730) of circa 4.7 kb, each encoding two ORFs arranged in an ambisense (but pointing outward) coding strategy. These two ORFs (hereafter ORFA and ORFB) were predicted to encode proteins that had no conserved homology with any protein in the viral database, except for two orphan proteins in the nr database from *Agaricus bisporus* (Table 1). Such proteins were encoded by an ORFan segment detected in a previous study, but not included by NCBI in the viral database (Deakin et al. 2017). By specific qRT-PCR, we assigned each of the two putative 4.7 kb virus segments to ORM fungi in the genus *Tulasnella* (Supplementary Table S5 and Supplementary online Fig. S7A). The specific contig DN30310 was abundant (ct 14 in qRT-PCR) in *Tulasnella* isolate O10, whereas contig DN33370 was less abundant in three *Tulasnella* isolates, O7, O11, and O12 (Supplementary online Table S5). A graphic representation of the main genomic features of these putative viral contigs is displayed in Fig. 6A. We named these two putative viruses Tulasnella ambivirus 1 and Tulasnella ambivirus 2 (TuAmV1 and TuAmV2, respectively). Given the surprising and unprecedented genomic organization of the segments assembled *in silico*, we checked by RT-PCR (overlapping fragments and nearly full length) if the RNA segment existed as a full-length sequence. RT-PCR revealed that indeed such segments existed as predicted *in silico* (Supplementary online Fig. S7B and C).

We then proceeded to investigate the presence of RNA transcripts (contigs) homologous to the TuAmV1 and TuAmV2 found in the XEO library in the other ERM and ORM fungal libraries described in this work. Homologous segments were absent in ERM fungi but were found in the other ORM fungal libraries. In detail, four putative full-length segments were identified and associated with specific isolates of *Tulasnella* and *Ceratobasidium* (Supplementary online Table S5), whereas two more segments were likely only partial genomes. We named these new viruses Ceratobasidium ambivirus 1 (CeAmV1=contig DN43545), Tulasnella ambivirus 3, 4, 5 (TuAmV3, TuAmV4, TuAmV5, respectively contigs DN37145, DN32762, and DN36670); the two partial genomes belong to Tulasnella ambivirus 6 and 7 (contigs DN36393 and DN36572). In all cases, the genomic organization of the contig assembled *in silico* is ambisense, but in one case the two ORFs point inward, as is common for other ambisense segments in negative stranded RNA viruses (Supplementary online Fig. S8). The Trinity-assembled contig DN43545 had a length of 9.8 kb, but a closer inspection revealed that the contig was an almost complete dimer (Supplementary online Fig. S10C) and that the monomeric sequences were similar, in length and genome organization, to the others retrieved.

The unusual ambisense genomic organization is also supported by the number of overlapping reads mapping across the full-length genome visualized with Tablet (Supplementary online Fig. S9). DNA corresponding to the RNA segment could not be detected by a sensitive qPCR assay in any of the isolates carrying these two virus genomic segments (Supplementary online Fig. S10A) therefore excluding i) that they are transcripts derived from a DNA genome, ii) that they are transcripts derived from endogenization of an RNA virus or iii) that their replication occurs through a DNA intermediate.

Given the number of different ambiviruses found in our collection of ORM fungi, all having a genome with RNA molecules between 4 and 5 kb in length, we could align the respective deduced encoded proteins (ORFA-like and ORFB-like) and revealed that they were both fairly conserved. ORFA encodes a protein with a well conserved domain centered around a GDD motif, as is the case with most of the RdRPs of RNA viruses (te Velthuis 2014) (Fig. 6B). Also, some other well conserved key residues of other RdRP domains were present, but insufficient to be detected as RdRP domains with the current motif search engines (MOTIF Search at https://www.genome.jp/tools/motif/). Overall, the estimate of the evolutionary divergence among aligned ORFA proteins was 0.66 (for specific pairs see Supplementary online Fig. S11A). ORFB-encoded proteins were also conserved among different isolates, but less than ORFA-encoded proteins (estimate of the average of evolutionary divergence is 0.72). Some residues were conserved among all isolates so far described, but this protein remains with no assigned function at this time (Supplementary online Fig. S11B). Given the limited conservation with other viral RdRP, a phylogenetic tree that includes them would not be reliable (Holmes & Duchene 2019).

The prevalence of this new putative viral genomic segment in many of our *Tulasnella* isolates prompted us to search for homologues in libraries we previously published from other filamentous fungi: one homologous viral fragment was retrieved from library F4 from the esca disease-associated fungal collection (Nerva et al. 2019a). The main features associated to this virus fragment are in Supplementary online Fig. S12.

The existence of closely related ambisense contigs having opposite orientation, inward and outward (Fig. 6 and Supplementary online Fig. S8), suggested the possibility that they are alternative assembly forms originating from a circular RNA molecule, or from a head-to-tail dimer molecule originated as a replication intermediate. We looked for evidence of such molecules by RT-PCR across a putative circularized junction (and by visualization of reads mapping across the putative junction). Both approaches (Supplementary online Fig. S9 and S10B and S10C) supported the possibility of a circular or dimer replication intermediate (in comparison with mapping of a traditional plus strand RNA segment).

Northern blot analysis with positive and negative sense probes corresponding to each of the two ORFs was carried out for two viruses, TuAmV1 and CeAmV1 (found in O10 and O4 isolates, respectively). Surprisingly, the two viruses yielded a different banding pattern. In the case of TuAmV1, the four probes reacted specifically with a single RNA of the expected genomic size (Fig. 7A). In the case of CeAmV1, the most abundant band has an estimated size consistent with a putative dimer of the genomic segment, which is instead present at lower concentration (Fig. 7B). We could not reveal the existence of subgenomic RNAs. A difference between TuAmV1 and CeAmV1 was also observed in

relation to positive sense and minus sense mapping reads: we arbitrarily assigned plus strand orientation to the RNA that encodes for ORFA, the putative RdRP, and minus strand to the opposite orientation; in the case of TuAmV1, we detected more plus strand reads, whereas the contrary was true for CeAmV1 and all the other ambiviruses (Supplementary online Table S6).
We then proceeded with an RNase R assay to check if these RNAs were sensitive to its exonuclease activity. A time course shows that the RNA was completely degraded 30 min after adding the enzyme, like a ToBRFV control we included in the experiment, thus suggesting that these RNA species might not be circular (Supplementary online Fig. S13).
Finally, a cell fractionation protocol was performed in order to check in which fraction the virus RNA was most enriched: viral RNA did not sediment after low speed centrifugations, but accumulated mostly in the microsomal fraction (in the pellet of a high speed centrifugation, after rounds of differential centrifugations) (Supplementary online Table S7). Despite its accumulation in the microsomal fraction, we were not able to see any specific virus-like morphology associated with the presence of the RNA in this fraction.

4.5 Discussion

The close and evolutionarily long-term association between mycorrhizal fungi and their host plants offers a wide temporal window for possible inter-kingdom horizontal virus transfers and selection of mycoviruses with a potential beneficial role in the tripartite interactions involving mycoviruses, mycorrhizal fungi and host plants. However, a prerequisite for testing these hypotheses is the comprehensive description of the viromes associated with mycorrhizal fungal isolates. In the present study, we describe the virome of a collection of ERM and ORM fungi that results in the discovery of surprising new evolutionary lineages that extend the boundaries of homology-based searches of RdRP conserved domains. Our work reports, for the first time, mycoviruses associated with ERM fungi and increases the diversity of mycoviruses characterized in ORM fungi (Ong et al. 2016; Ong et al. 2017; Ong et al. 2018). In addition to new viruses related to established taxa, we present three virus types with unprecedented genome organizations, one of which represents a completely new virus group that was not possible to reliably include in the monophyletic phylogenetic tree comprising the vast majority of RNA viruses (Wolf et al. 2018). The ICTV currently includes all viruses with an RNA genome in the realm *Riboviria*, and those that specifically use an RdRP for their replication are in the kingdom *Orthornavirae*. The five branches in the *Orthornavirae* monophyletic tree are now recognized as phyla: the *Lenarviricota* (levi-narna-ourmia-mito-like viruses), the *Pisuviricota* (the picornavirus supergroup), the *Kitrinoviricota* (including the alphavirus and flavivirus supergroup), the *Duplornaviricota* (including a number of dsRNA virus clades) and the *Negarnaviricota* (including viruses with (-)RNA virus genomes) (Koonin et al. 2020). The viruses from our current study that could be reliably placed in this monophyletic tree belong to the *Lenarviricota*, *Kitrinoviricota* and *Negarnaviricota*.

4.5.1 Discovery of a novel split organization of the palm domain in two genomic viral segments

The ERM fungi harbored two distinct narna-like contigs, DN37559 and DN43802, which showed some conservation with Plasmopara viticola lesion associated narnavirus 20 (unpublished, GenBank QIR30299.1) and Beihai narna-like virus 22 (Shi et al. 2016), respectively. However, a closer inspection of DN37559, DN43802 and Beihai narna-like virus 22 alignments revealed that the two lenar-like contigs hosted by ERM fungi aligned to different regions of the RdRP of Beihai narna-like virus 22. Unexpectedly, DN37559 harbored the A and B subdomain of the palm domain at the carboxy-terminus, whereas DN43802 contained the C and D subdomains at its amino terminus. A number of viruses with the same domain organization as that of DN37559 have already been detected *in silico* in previous works (Osaki et al. 2016; Zoll et al. 2018; Nerva et al. 2019a), but only in a recent report was the absence of the important GDD carrying C subdomain noticed (Lin et al. 2020). For most of these lenar-like viruses, neither the raw reads nor the transcriptome assembly were deposited in public databases, and it was therefore impossible to verify a hypothesis of strict association of DN37559- and DN43802-type segments, with the exception of a library of fungi associated with esca disease with reads deposited in SRA archives (Nerva et al. 2019a), where a homologue of DN43802 was indeed found. Furthermore, we found in transcriptomic libraries from our previous works the same combination of DN37559-type and DN43802-type homologous segments showing that these viruses are not specific to *O. maius*, but can be found in both asco- and basidiomycetes.

A comprehensive phylogeny of RNA viruses is based on the viral RdRP universally conserved module, which includes at least the A, B and C subdomains. Here we show for the first time that such subdomains can be encoded by two distinct genomic segments: subdomain A and B from an DN37559-type of virus segment and C from an DN43802-type of segment. Previously, an exception to the A-B-C subdomain module was reported, showing a circular permutation of the order of the domains (C-A-B) in *Permutotetraviridae* and *Birnaviridae* families (Gorbalenya et al. 2002). Nevertheless, also in this case, the palm domain was encoded by a single protein. We can safely exclude that splitting of the coding sequence in two contigs is due to an assembly or sequencing artefact because i) we could confirm the size of each segment by Northern blot, ii) we could complete a RACE experiment with both viruses, iii) the read coverage is very high, and iv) the same single contigs are reported independently from different works. The highly conserved sequences in the 5’ and 3’ UTR support the notion that the two sequences constitute the genome of a single, bisegmented virus for which we suggest the name Oidiodendron maius splipalmivirus 1 (OmSPV1). Splipalmiviricetes/splipalmiviridae is the tentative name of a taxon (likely a new class or new family) that accommodates this new clade of viruses (name derived from SPLIt PALM domain viruses). Bi-partitism in some narnaviruses was first inferred from metagenomic studies (Shi et al. 2016) and was recently confirmed to be a feature of some protozoa and apicomplexan-infecting viruses (Charon et al. 2019). Nevertheless, in such cases the RdRP palm domain is present in a single protein encoded by one of the two viral genomic segments, while the other segment codes for proteins of unknown function.

This observation can be extended to the thousands of RNA viruses so far characterized. To our knowledge, there is no precedent for a splitting of the RdRP into two distinct proteins. This raises a question about the possible evolutionary trajectory of these viruses. Judging from their very distant relationship from existing RNA viruses, it might be tempting to hypothesize that they are basal to narnaviruses. Alternatively, it is also possible that having the palm subdomains encoded by different proteins might alleviate some of the structural constraints that are present in the single protein-encoded ABC palm subdomain, allowing faster evolution after separation of the domains.

Another interesting feature of these viruses is the fact that, for both segments, there is a higher accumulation of the (-)RNA (arbitrarily assigning the plus strand to the one encoding the RdRP, as for other narnaviruses). This is a new feature (confirmed by strand-specific Northern blot analysis) that has never been reported before in any other putative (+)RNA. Viruses are defined as carrying a plus or minus strand genome based on the fact that encapsidated RNA respectively can or cannot function as messenger RNA once they are released in the cell: given that narnaviruses are capsidless naked RNA virus-like agents, such definition does not apply.

Narnaviruses were first discovered in baker's yeast, where it was not possible to associate any phenotypic change with their presence-absence (Kadowaki & Halvorson 1971; Wesolowski & Wickner 1984; Hillman & Cai 2013). The recent evidence that narnaviruses can infect not only fungi but also insects (specifically dipteran) (Chandler et al. 2015; Goertz et al. 2019), other arthropods (Shi et al. 2016), a brown alga (Waldron et al. 2018), trypanosomatids (Grybchuk et al. 2018) and nematodes (Richaud et al. 2019) has raised further interest in these viruses. Moreover, a recent work has shown their effect on the reproductive fitness of their Mucoromycota host *Rhizopus microsporus* in a tripartite interaction with an endobacterium that is involved in virulence toward the plant host (Espino-Vazquez et al. 2020). The ERM strains hosting OmSPV1 have been isolated from *C. vulgaris* from a heathland area in Northern Italy and are genetically close but distinct (Perotto et al. 1996). Indeed, the success of ericaceous shrubs in heathland habitats, which can be very poor in nutrients and high in toxic metal ions, is due to their endomycorrhizal association (Bradley et al. 1982). The finding of viral sequences in such strains poses the question of whether they might play a role in the fungus-plant, fungus-environment and plant-environment interaction, in analogy with the thermal tolerance observed for a tripartite interaction plant-fungal endophyte-virus (Márquez et al. 2007).

In ERM fungi, we also found a new species of ourmia-like virus, with characteristics similar to ourmia-like mycoviruses already characterized in other fungi. This group of viruses has gained some attention because it shows evidence of evolutionary exchanges between different kingdoms (Rastgou et al. 2009). Fungal isolates E35 and E36 (hosting OmOIV1) originate from distinct soil plots within the same polluted area (Lacourt et al. 2000; Martino et al. 2003, Vallino et al. 2011) and have been found to tolerate heavy metals (Zn and Cd, respectively) (Vallino et al. 2011). An ourmia-like virus was shown to be associated with mitochondria (Hrabáková et al. 2017) and this fact can provide a possible functional link between virus presence and heavy metal tolerance, since mitochondria are central to heavy metal tolerance in fungi (Daghino et al. 2019). The possibility to transfect

O. maius protoplasts with transcripts from a putative infectious clone will help determine if OmOIV1 indeed bears a role in heavy metal tolerance.

4.5.2 An unprecedented tri-cistronic RNA related to members of *Bunyavirales* encodes both a putative RdRP and a putative Nc on the same viral segment

The virome of ORM fungi has been a subject of a pioneering study that revealed the presence of dsRNA that did not result in a specific taxonomic assignment (James et al. 1998). Investigations carried out on natural populations of orchids from Australia revealed i) several new endornavirus species and a new endornavirus genomic organization (Turina et al. 2018), ii) ten partitiviruses from *Ceratobasidium* spp. (Ong et al. 2017), and iii) a mitovirus, a hypovirus, two dsRNA viruses, a toti-like virus and a megabirna-like virus also from *Ceratobasidium* spp. (Ong et al. 2018).

To this already vast array of viral sequences, we add here a new mitovirus, a new barnavirus and two new endornaviruses, all related to those already found but different enough to be considered new species. In addition, for the first-time a minus stranded RNA virus, TuBlV1, was detected in ORM fungi. Phylogenetic analysis of the RdRP encoded by this virus shows that it belongs to the *Bunyavirales* order, but its genome has some unique features that expand the diversity of genome organizations so far characterized in this order. We demonstrate the existence of a single genomic segment that also encodes, together with the RdRP, another small ORF in the ambisense orientation (genomic sense) and a third ORF that terminates just upstream of the RdRP, in the same antisense orientation and with no intergenic region.

In the *Arenaviridae*, only the *Mammarenavirus* and *Reptarenavirus* genera have an RdRP encoding segment that also encodes for a second ORF in the ambisense orientation, a small matrix protein (Perez et al. 2003), and this is so far a unique feature inside the hundreds of viruses characterized in the *Bunyavirales* (Shi et al. 2016; Kaefer et al. 2019). Nevertheless, we show evidence that in the case of TuBlV1 the protein encoded in the ambisense orientation is not a matrix protein but a nucleocapsid protein, whereas in the case of the bipartite arenaviruses the nucleoprotein is encoded by a different genomic segment, as is the case in all previously known members of the *Bunyavirales*. Also puzzling is the presence of the third ORF upstream of the RdRP. We confirmed this occurrence in a contiguous RNA both by sequencing the RT-PCR product across the junction and by Northern blot analysis. A careful look at the reads mapping across the junction also confirms the uniformity of the sequence in that region without any detected variants. The last nucleotide encoding ORF3 is the first nucleotide of the RdRP, opening the possibility of a -1 frameshift as an expression strategy of the RdRP encoding ORF or of a re-initiation expression strategy.

Phylogenetic analysis also shows that this virus does not belong to any of the existing or proposed clades in the *Bunyavirales*, including also a number of recently characterized phlebo-like mycoviruses (Nerva et al. 2019b; Chiapello et al. 2020); therefore a new taxon should be assigned to accommodate TuBlV1.

4.5.3 Orchid mycorrhizal fungi harbor ambiviruses, an ORFan group of ambisense bicistronic viruses that cannot yet be assigned to the viral RdRP monophyletic phylogenetic tree.

In our collection of ORM fungi, we serendipitously found a number of distinct virus sequences, characterized by a bicistronic ambisense RNA segment of circa 5 kb with both ORFs conserved among the discovered virus segments; these features define a new virus clade with no detectable relationship with existing characterized viruses. Nevertheless, we could show that they are replicated via a minus strand/plus strand RNA replication cycle that does not entail a DNA phase. The surprising ambisense orientation of the ORFs encoded by these genomes have so far only been shown for the *Bunyavirales* and more recently for some narna-like sequences (DeRisi et al. 2019; Dinan et al. 2020), although in this case the evidence is indirect, and the two ORFs overlap. We could not show evidence of subgenomic RNA accumulation, but relatively abundant accumulation of both plus strand and minus strand orientation of the genomic RNA could suggest that the genomic and antigenomic RNA can express both proteins. Furthermore, both proteins (ORFA- and ORFB-derived) are highly conserved among the different isolates, suggesting a function related to viral replication. Since the assembly of ambivirus contigs by Trinity software was in some cases prone to some artifacts, leading to arbitrary assembling of contigs as dimers with a number of reads across the putative junction, we deposited the putative ambivirus genomes as monomeric sequences. Our results indicate that abundant accumulation of the dimer is not a prerequisite of the replication cycle because one of these ambiviruses (TuAmV1) does not accumulate the dimer as shown by Northern blot. Abundant genomic dimer accumulation was previously shown for some specific bunyavirus-host combinations (Bertran et al. 2016), and their significance as defective interfering RNA that regulate accumulation of the genomic RNA was hypothesized. A likely misassembly due to the continuum of virus reads mapping across a dimer also occurred when a contig expressing an ORF (with homology to ORFB of ambivirus) was first found in *Agaricus bisporus* (ORFan 1, GenBank ID KY357519) (Deakin et al. 2017). The ORFan 1 includes five ORFs in addition to the one showing homology to ORFB of ambivirus and ORFA is fragmented, possibly due to contig misassembly. The authors found ORFan 1 in two of their isolates at very low transcript accumulation, and thus, ORFan 1 was deposited in GenBank as an *Agaricus bisporus* sequence, therefore not included in any viral database. Here, we provide for the first time evidence that homologues of this ORFan are viral genomic fragments encoding a putative RdRP.

In fact, even though a motif scan search of the two ORFs encoded by the ambivirus did not detect any conserved motif, a closer inspection of conserved motifs in protein alignments of ORFAs suggests the possible existence of an uncanonical RdRP palm domain. As discussed above, motifs A, B and C subdomains are crucial for enzyme activity. To find evidence suggesting the RdRP nature of ORFA from ambiviruses, we aligned ORFAs from all the ambiviruses discovered in the work with a group of different RdRPs hosting the canonical disposition previously used for detecting motifs A, B and C permutation (Gorbalenya et al. 2002). Results show that motif C (hosting the hallmark motif GDD) is aligned correctly with the GDD triad from ambiviruses, while a partial match is found for motifs A and B. Subdomain A usually hosts a conserved D residue with

another D residue separated by four or five amino acids (D-X5-4-D); the second of these D residues is not conserved in *Mononegavirales* and *Bunyavirales* (te Velthuis 2014). ORFA proteins encoded by ambiviruses show only the first conserved D residue and the F residue in position +4, which is the most common amino acid in +strand RNA viruses therefore mixing features of both plus strand and minus strand RdRP in this subdomain (te Velthuis 2014). Motif B shows three conserved residues: G, T and N in plus strand RNA viruses, separated by non-conserved amino acids (G-X2-3-T-X3-N). In this case, a partial similarity to ambivirus ORFA could be found with the conserved G residue. However, instead of T and N, the conserved G residue of ORFA was followed by a second conserved G residue in position +3, and N and ST conserved residues downstream.
In conclusion, partial conservation could be observed between the active site of viral RdRPs and conserved motifs of ambivirus ORFA products, with A and C putative subdomains more closely related to +strand (and -strand) RNA viruses, while subdomain B is less recognizable, with a number of conserved residues that match poorly to subdomain B from other viruses. Despite this variability and the chimeric nature of the ABC subdomains in the ambivirus, the three amino acid residues that are invariant (D in subdomain A, G in subdomain B and D in subdomain C) are indeed conserved: in particular, the two aspartic acid residues can interact with two Mg++ ions in the catalytic core of the domain, whereas the glycine in subdomain B is necessary for nucleotide selection (Steitz 1998). Attempts at purifying virus particles or nucleocapsids associated with this virus segment failed, but the viral RNA enriches in the microsomal fraction, as it is the case of other capsidless viral elements (Jacob-Wilk et al. 2006).

4.6 Conclusions

The virome in our collection of ERM and ORM fungi featured known mycoviruses as well as novel viruses not previously described in fungi. The identification of new viruses in mycorrhizal fungi expands the boundaries of characterized RNA virus diversity and raises the question of whether mycorrhizal fungi may represent a special ecological niche for these novel viruses. A more extensive search in fungi with different lifestyles will clarify this point.

4.7 Acknowledgments

We thank Riccardo Lenzi, Andrea Delliri, and Caterina Perrone for technical assistance and Prof. Bryce Falk (UC-DAVIS, USA) for critical discussion of the results. Suvi Sutela and Eeva Vainio were supported by the Academy of Finland (grant number 309896). Marco Forgia was supported by a Ph.D. fellowship by MIUR.

4.8 Figures and Tables

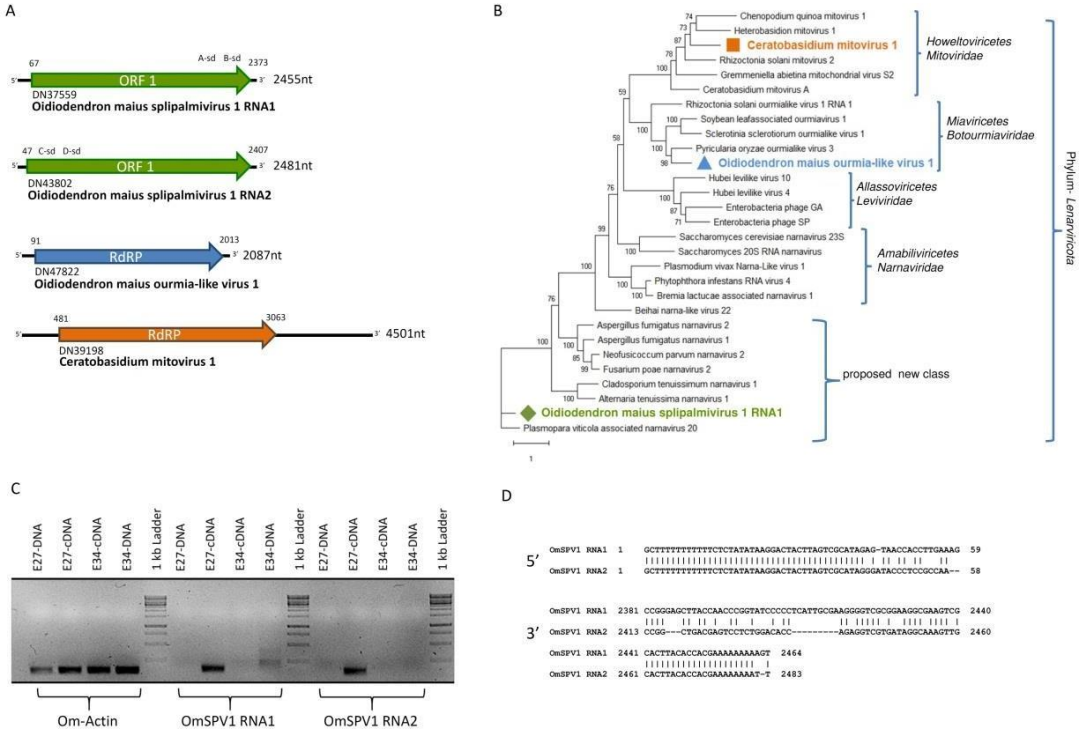


Figure 1. Main features of contigs related to members of *Lenarviricota* from ERM and ORM fungi. A: Schematic representation of genome organization, with the main features. RdRP=RNA-dependent RNA polymerase; ORF=open reading frame; nt=nucleotides; A-sd, B-sd, C-sd and D-sd are, respectively the A, B, C and D subdomains of the RdRP palm domain. B: Phylogenetic tree derived from alignments of the most closely related putative RdRP amino acid sequences inside the *Lenarviricota* clade. Model of substitution: Blosum62+F+I+G4. Consensus tree is constructed from 1000 bootstrap trees. Log-likelihood of consensus tree: -49081.578022. At nodes, the percentage bootstrap values. The main existing and proposed taxonomic clades (class and family for the five smaller parentheses, phylum for the large one) are grouped with parenthesis. C: TAE 1.5% agarose gel of qPCR products to show that there is no DNA template corresponding to transcripts of Oidiodendron maius splipalmivirus 1 RNA1 and RNA2 (OmSPV1 RNA1 and OmSPV1 RNA2). As a control for amplification we included a fragment of the *O. maius* actin gene. Lanes are labeled with the template used for the qPCR reaction and include an infected *O. maius* isolate (E27) and an uninfected isolate (E34). D: 5' and 3' untranslated region of the sequence alignment of OmSPV1 RNA1 and OmSPV1 RNA2 genomic fragments.

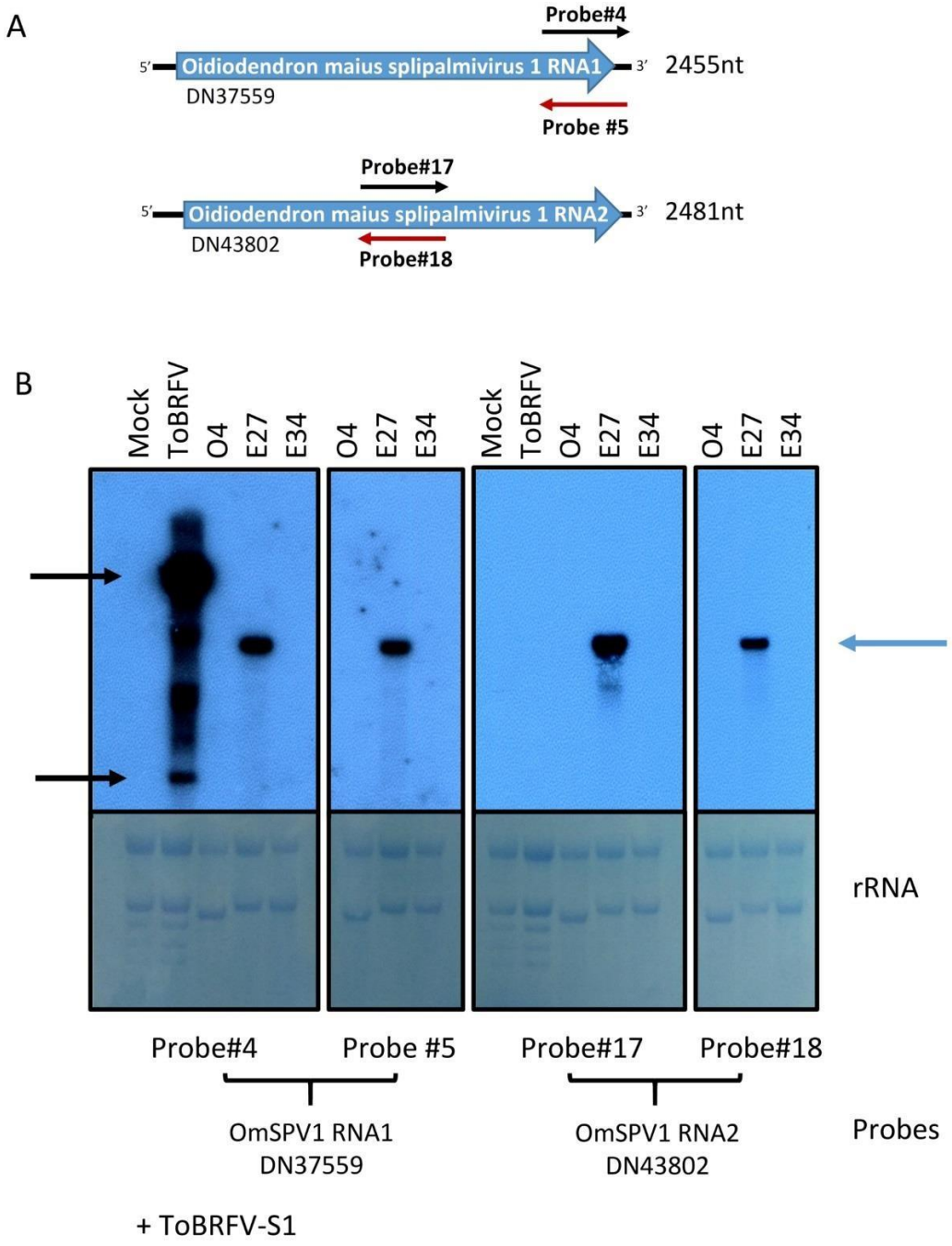

A
Probe#4
5′
Oidiodendron maius splipalmivirus 1 RNA1
3′
2455nt
DN37559
Probe #5
Probe#17
5′
Oidiodendron maius splipalmivirus 1 RNA2
3′
2481nt
DN43802
Probe#18
B
Mock
ToBRFV
O4
E27
E34
O4
E27
E34
Mock
ToBRFV
O4
E27
E34
O4
E27
E34
rRNA
Probe#4
Probe #5
Probe#17
Probe#18
OmSPV1 RNA1
DN37559
OmSPV1 RNA2
DN43802
Probes
+ ToBRFV-S1

Figure 2. Northern blot analysis of narna-like contigs DN37559 and DN43802 representing the two genome segments of the Oidiodendron maius splipalmivirus 1 (OmSPV1). A: Schematic representation of the position of the run-off transcript probes with codes identifying their orientation. In black, sense-oriented transcripts that hybridize with minus sense anti-genomic RNA intermediate. In red, antisense-oriented transcripts that hybridize with plus sense genomic RNA. **B:** Top panels, autoradiography exposed 2 hrs with samples of total RNA (circa 3 ug/gel lane). The RNAs in the panel on the far left were hybridized first with probe #4 and subsequently with a tomato brown rugose fruit probe (ToBRFV-S1) that can can be used as standard for RNA size (6.3 Kb for the genomic RNA and 0.7 kb for subgenomic RNA2). Lower panel is a methylene blue stained membrane to show different ribosomal RNA loadings (rRNA). Sample nomenclatures include E27 as the infected *O. maius* isolate and isolates O4 and E34 as negative controls. Mock is RNA extracted from a mock inoculated tomato plant. A blue arrow in this panel points to the position of the OmSP1 genomic RNAs, whereas black arrows points to the genomic and subgenomic RNA2 of ToBRFV.

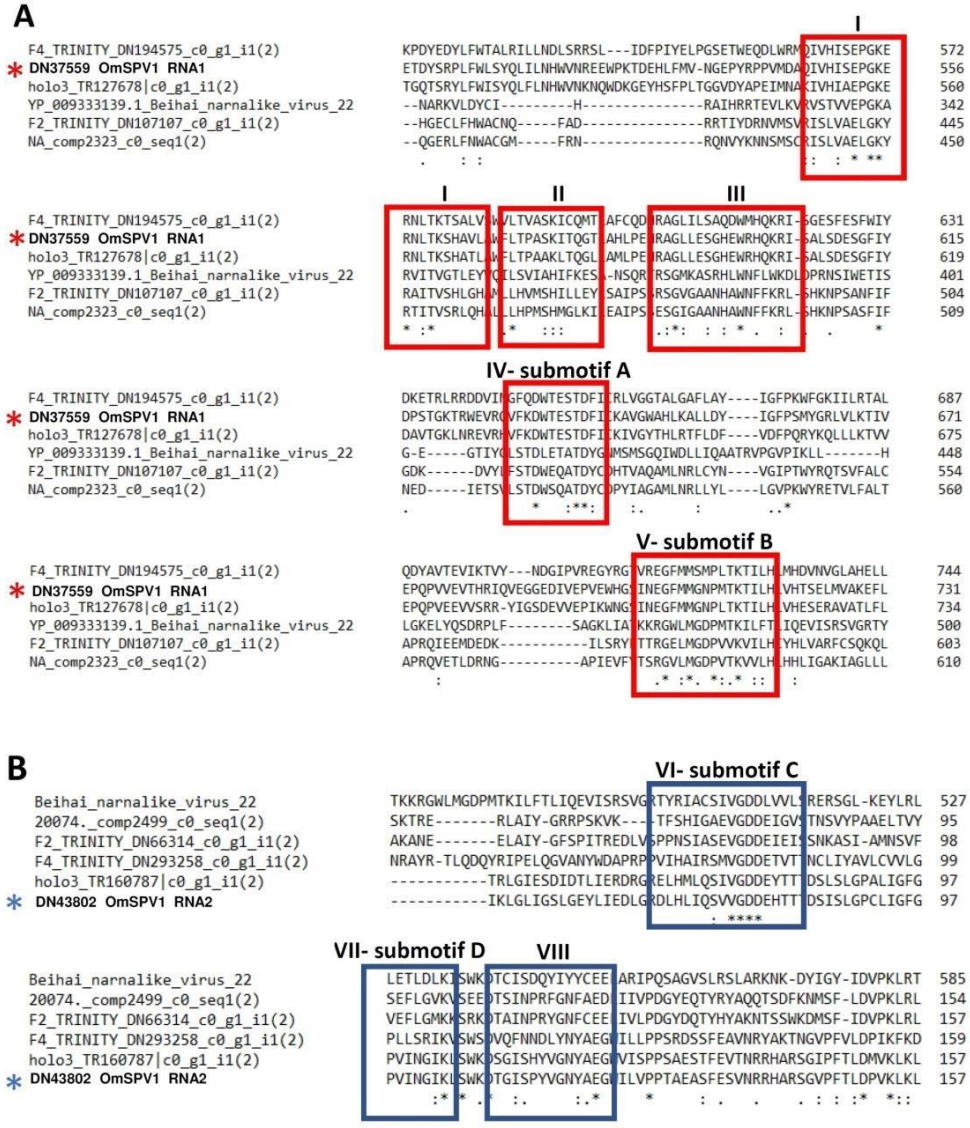
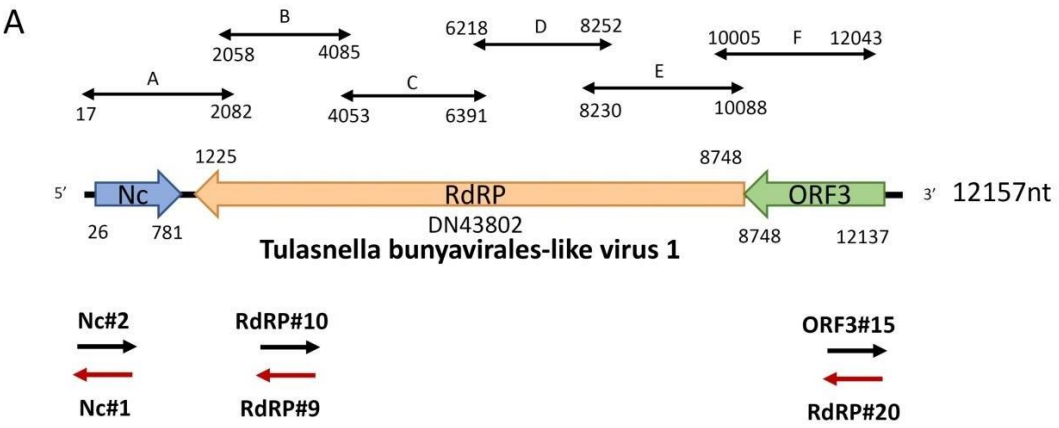


Figure 3. Protein alignments of contigs DN37559 and DN43802 representing the two genome segments of the Oidiodendron maius splipalmivirus 1 (OmSPV1). A: Five protein sequences retrieved from different previously published assemblies from fungal RNAseq projects with some homology to OmSPV1 RNA1. Only the carboxy terminal of the proteins is shown. **B**: Proteins from the same libraries with homology to OmSPV1 RNA2 ORF encoded protein. Only the amino terminal of the protein alignment is displayed. OmSPV1 RNA1 is marked with a red asterisk and OmSPV1 RNA2 is marked with a blue asterisk. In the alignment, eight conserved regions already characterized for viral RdRPs

could be observed and are marked with red rectangles for OmSPV1 RNA1 -related viruses and blue rectangles for OmSPV1 RNA2 -related sequences. Conserved regions are named I to VIII following Koonin et al. (Koonin 1991); regions IV, V, VI and VII are linked to motifs A, B, C and D from the RdRP palm subdomain, respectively. 20074 is the code for the fungal isolate MUT4935 isolated from *Posidonia oceanica*. F2 and F4 are libraries from mixed fungal isolates from esca-infected grapevines, and Holo3 is a library that comprises a number of fungal isolates from *Holothuria polii*.


A
B
2058
4085
6218
D
8252
10005
F
12043
A
17
2082
C
4053
6391
E
8230
10088
1225
8748
5′
Nc
RdRP
ORF3
3′
12157nt
26
781
DN43802
8748
12137
Tulasnella bunyavirales-like virus 1
Nc#2
RdRP#10
ORF3#15
Nc#1
RdRP#9
RdRP#20

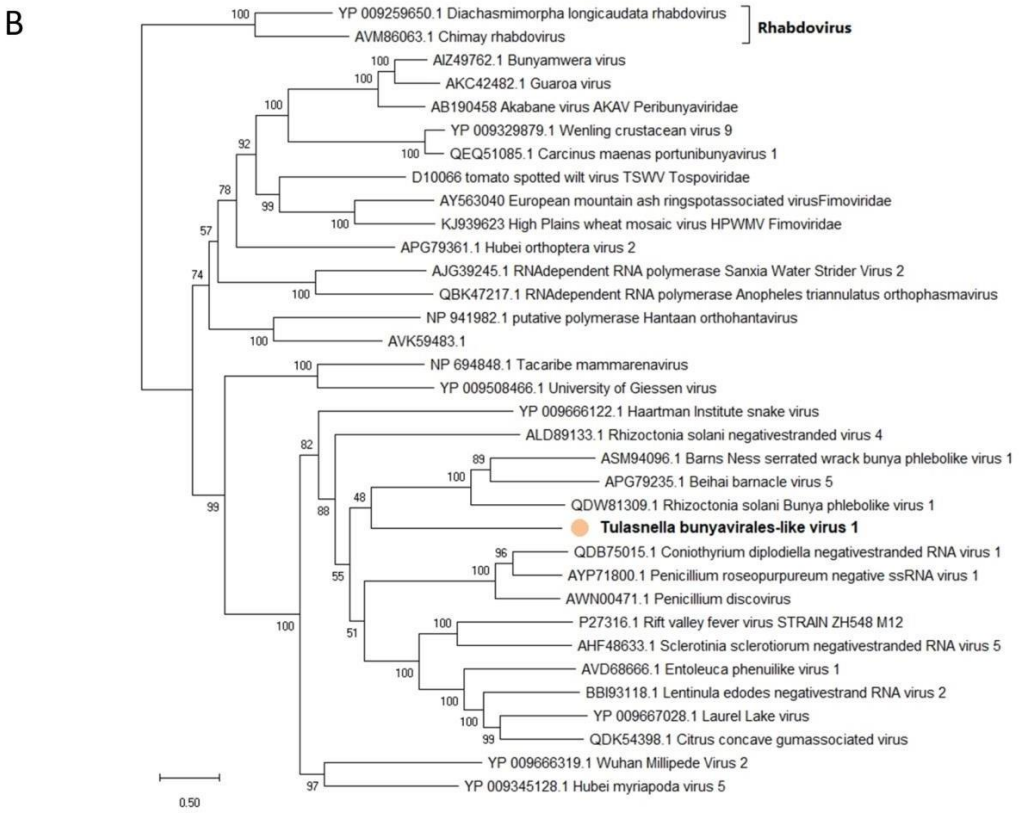

B
YP 009259650.1 Diachasmimorpha longicaudata rhabdovirus
AVM86063.1 Chimay rhabdovirus
Rhabdovirus
AIZ49762.1 Bunyamwera virus
AKC42482.1 Guaroa virus
AB190458 Akabane virus AKAV Peribunyaviridae
YP 009329879.1 Wenling crustacean virus 9
QEQ51085.1 Carcinus maenas portunibunyavirus 1
D10066 tomato spotted wilt virus TSWV Tospoviridae
AY563040 European mountain ash ringspotassociated virusFimoviridae
KJ939623 High Plains wheat mosaic virus HPWMV Fimoviridae
APG79361.1 Hubei orthoptera virus 2
AJG39245.1 RNAdependent RNA polymerase Sanxia Water Strider Virus 2
QBK47217.1 RNAdependent RNA polymerase Anopheles triannulatus orthophasmavirus
NP 941982.1 putative polymerase Hantaan orthohantavirus
AVK59483.1
NP 694848.1 Tacaribe mammarenavirus
YP 009508466.1 University of Giessen virus
YP 009666122.1 Haartman Institute snake virus
ALD89133.1 Rhizoctonia solani negativestranded virus 4
ASM94096.1 Barns Ness serrated wrack bunya phlebolike virus 1
APG79235.1 Beihai barnacle virus 5
QDW81309.1 Rhizoctonia solani Bunya phlebolike virus 1
Tulasnella bunyavirales-like virus 1
QDB75015.1 Coniothyrium diplodiella negativestranded RNA virus 1
AYP71800.1 Penicillium roseopurpureum negative ssRNA virus 1
AWN00471.1 Penicillium discovirus
P27316.1 Rift valley fever virus STRAIN ZH548 M12
AHF48633.1 Sclerotinia sclerotiorum negativestranded RNA virus 5
AVD68666.1 Entoleuca phenuilike virus 1
BBI93118.1 Lentinula edodes negativestrand RNA virus 2
YP 009667028.1 Laurel Lake virus
QDK54398.1 Citrus concave gumassociated virus
YP 009666319.1 Wuhan Millipede Virus 2
YP 009345128.1 Hubei myriapoda virus 5
0.50

Figure 4. Main features of the Tulasnella bunyavirales-like virus 1 (TuBlV1). A: The genome organization of TuBlV1 with the main genomic features (ORFs and domains). Black bi-directional arrows indicate the position of RT-PCR amplification products to cover the whole genome. Unidirectional black and red arrows represent the position and orientation of the run-off transcripts used as probes in Northern blot analysis. Nc=Nucleocapsid; RdRP=RNA-dependent RNA polymerase; ORF=open reading frame; nt=nucleotide. **B**: Maximum likelihood phylogenetic tree derived from RdRP alignment of TuBlV1 with a number of bunyavirales representative of the main families in the order, (and two rhabdovirus used as outgroup). Amino acid substitution model is VT+F+I+G4. Log-likelihood of the tree: -200570.9602. Bootstrap values in percentage are displayed at each node. The tree is unrooted.

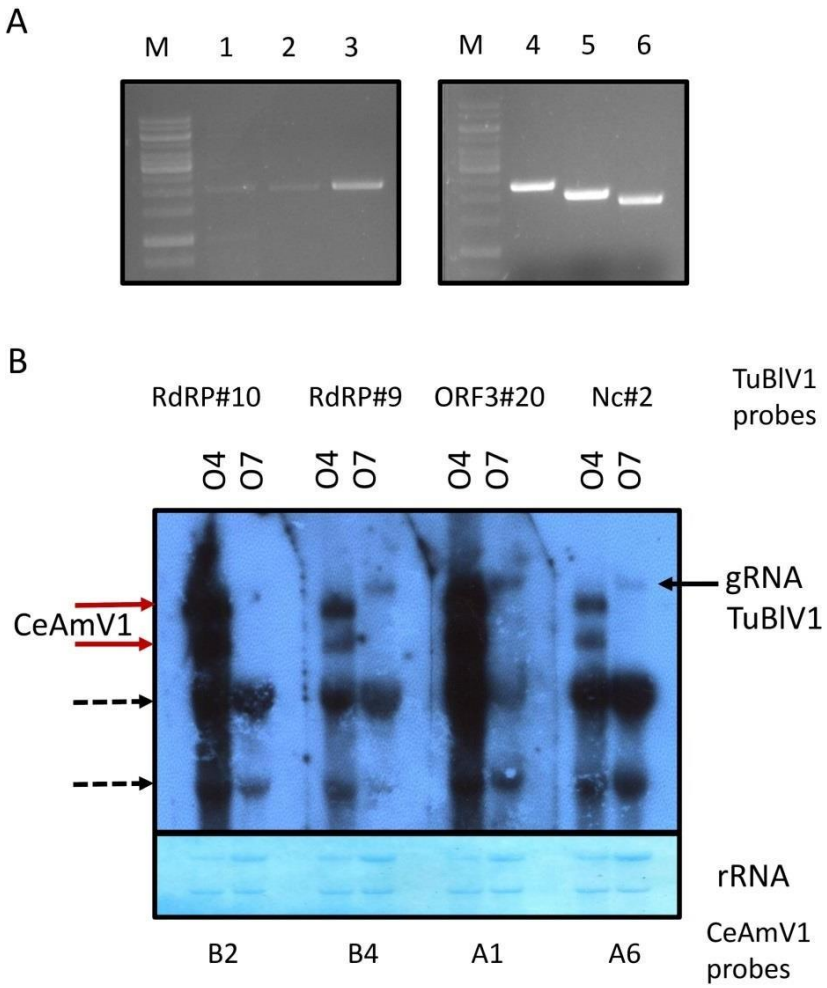


Figure 5. Evidence of a contiguous 12 kb genomic segment for Tulasnella bunyavirales-like virus 1 (TuBlV1). A: Overlapping RT-PCR of segments spanning the whole TuBlV1 genome. Lanes 1 to 6 correspond to segments A, B, F, C, D and E of Figure 4. M is the 1 kb Ladder. The three stronger bands correspond to 1 kb, 3 kb and 7 kb, respectively. **B:** Northern blot analysis of total RNA extracts from ORM fungal isolates O4 and O7 positive for Ceratobasidium ambivirus 1 (CeAmV1) and for TuBlV1, respectively. Specific probes for each of the two viruses were hybridized in succession, in order to derive the specificity of each probe. Here the result after the second hybridization is displayed, which therefore shows all the bands hybridizing with both probes. Red arrows point to the two bands specific for the ambivirus. Dotted black arrows point to the position of cross reactivity with ribosomal RNA. The black arrow points at the position of the TuBlV1 virus-specific RNA bands; upper panels are autoradiography exposed for 7 days. Lower panel is ribosomal RNA stained with methylene blue (the membrane picture was stretched vertically). Probes used are specified in Fig. 4 and Fig. 8; rRNA=ribosomal RNA.

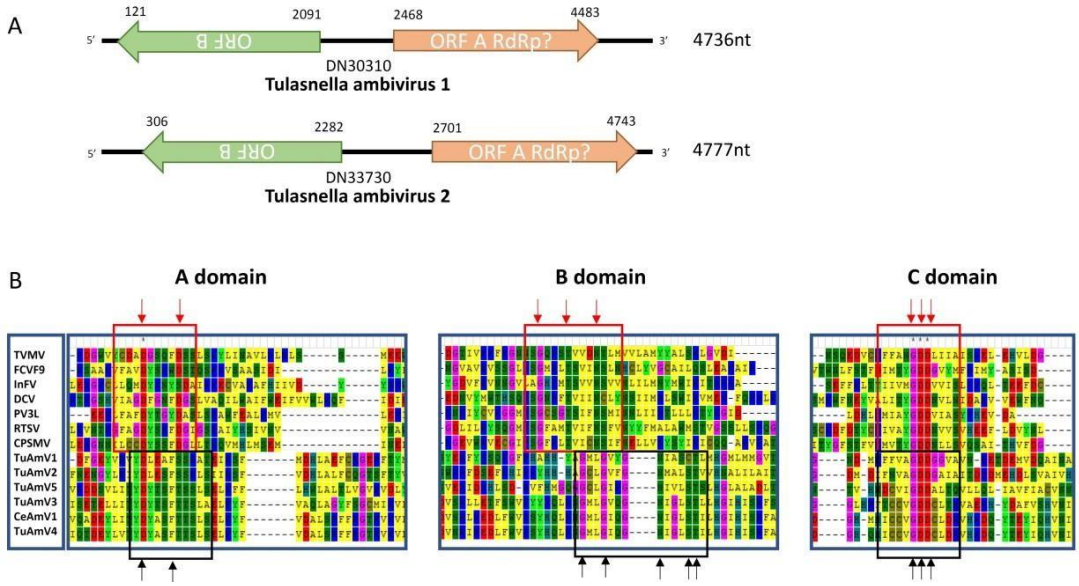


Figure 6. Genome organization of ambiviruses and their putative RdRP palm subdomains. A: Schematic representation of the genome organization of Tulasnella ambivirus 1 and Tulasnella ambivirus 2 (TuAmV1 and TuAmV2). Open reading frames (ORF) are represented by green (ORFB) and orange (ORFA) arrows. **B:** Alignment of conserved motifs of canonical viral RNA-dependent RNA polymerases (RdRPs) and ORFA proteins of ambiviruses. RdRPs were retrieved from alignment by Gorbalenya et al. (2002) and aligned using Clustal Omega. Conserved domains A, B and C were selected from the alignment and sequences were aligned again on Clustal adding the ORFA sequences of ambiviruses discovered in the present study. Results were displayed on MEGA6. Canonical motifs are surrounded by the red rectangles with conserved residues marked by the red arrows. The putative motifs and conserved amino acids of ambiviruses surrounded by the black squares and black arrows indicate conserved residues. The sequences used for the alignment are as follows: tobacco vein mottling virus (TVMV, 8247947); feline calicivirus F9 (FCVF9, 130538); infectious flacherie virus (InFV, 3025415); Drosophila C virus (DCV, 2388673); human poliovirus type 3 Leon strain (PV3L, 130503); rice tungro spherical virus (RTSV, 9627951); cowpea severe mosaic virus (CPSMV, 549316); Tulasnella ambivirus 1 (TuAmV1, MN793991); Tulasnella ambivirus 2 (TuAmV2, MN793992); Tulasnella ambivirus 5 (TuAmV5, MN793996); Tulasnella ambivirus 3 (TuAmV3, MN793994); Ceratobasidium ambivirus 1 (CeAmV1, MN793993); and Tulasnella ambivirus 4 (TuAmV4, MN793994).

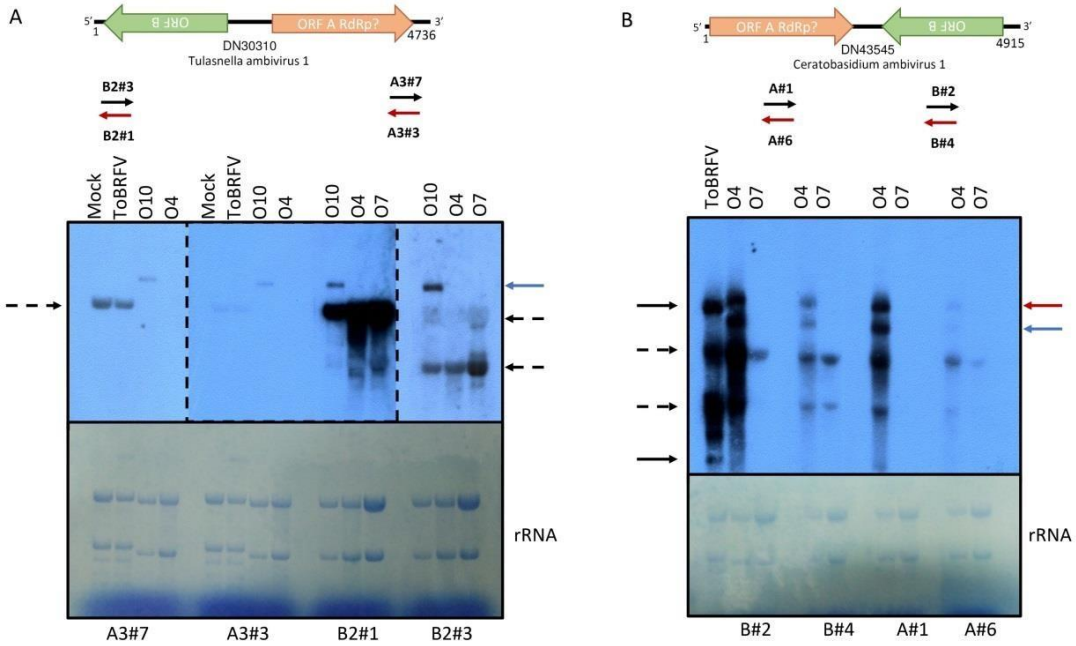


Figure 7. Northern blot analysis of total RNA extracted from ORM fungal strains harboring ambiviruses. A: ORM fungal strains O4, O7 and O10, of which the latter was infected by Tulasnella ambivirus 1 (TuAmV1). The probes used for TuAmV1 are presented at the bottom of each panel. Film was exposed to the membrane for 24 hrs. The blue arrow points to the position of the single specific band hybridizing with the probe. Dotted black arrows point to unspecific hybridization to ribosomal RNAs (rRNAs). **B:** ORM fungal strains O4 and O7 of which the former hosts Ceratobasidium ambivirus 1 (CeAmV1). Film was exposed to the membrane for 12 hrs. The left end panel was hybridized first with probe B2 and subsequently with a tomato brown rugose fruit virus probe (ToBRFV-S1) that can be used as standard for RNA size (6.3 Kb for the genomic RNA and 0.7 kb for subgenomic RNA2). In this panel a black arrow points to the position of ToBRFV genomic RNA and subgenomic RNA2, a blue arrow to the putative CeAmV1 genomic RNA and the red arrow to the position of the putative CeAmV1 dimer. Dotted black arrows point to the position of cross hybridizing rRNA. In both A and B, lower panels are methylene blue stained membranes rRNA loadings. Mock is RNA extracted from a mock inoculated tomato plant. At the top of each panel is a schematic representation of the position of the run-off transcript probes with codes that identify their orientation. In black, sense-oriented transcripts that hybridize with minus sense anti-genomic RNA intermediate. In red, antisense-oriented transcripts that hybridize with plus sense genomic RNA.

Table 1. Viral contigs found in our collection of ericoid and orchid mycorrhizal fungi. The main features of each virus contig are reported together with the suggested virus names. The first hit in our viral database BLAST search is also reported, with the main features of the alignment.

GenBank ID	Virus name	Domain[a]	Length	Contig[b]	Accession[c]	Name[c]	Per. Ident. [c]	Query cov [c]	E-value [c]	Mapping reads [d]
MN736964	Oidiodendron maius splipalmivirus 1 RNA1	N/A	2455	DN37559_c1_g1_i3	QIR30299.1	Plasmopara viticola associated narnavirus 20	48.519	86%	0	231825
MN736965	Oidiodendron maius splipalmivirus 1 RNA2	N/A	2481	DN43802_c0_g4_i1	APG76982.1	Beihai narna-like virus 22	26.37	23%	0.089	194000
MN736966	Oidiodendron maius ourmia-like virus 1	RdRP	2087	DN47822_c2_g3_i2	QAB47442.1	Combu positive-strand RNA mycovirus	60.741	77%	0	749959
MN736968	Ceratobasidium mitovirus 1	RdRP	4501	DN39198_c0_g3_i1	AWY10986.1	Sclerotinia sclerotiorum mitovirus 28	50.435	14%	2.24E-56	143938
MN738550	Ceratobasidium endornavirus 1	RdRP	21599	DN44545_c0_g1_i1	YP_009552276.1	Rhizoctonia solani endornavirus 1	40.344	72%	0	17370
MN738551	Ceratobasidium endornavirus 2	RdRP	18941	DN44360_c0_g1_i1	QDW65431.1	Rhizoctonia solani endornavirus 4	33.979	70%	0	~20000
MN738552	Tulasnella barnavirus 1	RdRP	4389	DN27349_c0_g2_i1	ALD89112.1	Rhizoctonia solani barnavirus 1	52.688	31%	3.70E-156	17296
MN793991	Tulasnella ambivirus 1	N/A	4724	DN30310_c0_g1_i1	AQM32755.1	hypothetical protein [Agaricus bisporus]	28.08	8%	0.051	5172
MN793992	Tulasnella ambivirus 2	N/A	4777	DN33730_c1_g1_i1	AQM32755.1	hypothetical protein [Agaricus bisporus]	31.46	5%	0.065	1662
MN793993	Ceratobasidium ambivirus 1	N/A	4915	DN43545_c0_g4_i1	AQM32755.1	hypothetical protein [Agaricus bisporus]	29.88	19%	4.00E-19	51 000
MN793994	Tulasnella ambivirus 3	N/A	5120	DN37145_c0_g1_i6	AQM32755.1	hypothetical protein [Agaricus bisporus]	30.35	17%	1.00E-24	1074
MN793995	Tulasnella ambivirus 4	N/A	4924	DN32762_c1_g1_i1	AQM32755.1	hypothetical protein [Agaricus bisporus]	30.49	19%	3.00E-22	620
MN793996	Tulasnella ambivirus 5	N/A	4632	DN36360_c0_g1_i2	AQM32754.1	hypothetical protein [Agaricus bisporus]	29.81	39%	2.00E-74	736
MN793997	Tulasnella bunyavirales-like virus 1	RdRP	12157	DN37204_c0_g2_i1	APG79361.1	Hubei orthoptera virus 2	26.891	5%	4.75E-07	13972

[a] Domains were searched using Motif Search at https://www.genome.jp/tools/motif/ using both Pfam and NCBI-CDD databases

[b] The code is that provided automatically by Trinity software after assembly. The suffixes in the contig IDs were removed in the text for simplicity.

[c] All the parameters refer to the first hit in BlastX searches of nr NCBI databases.

[d] Total number of reads present in each library mapping to each of the putative viral contigs assembled by Trinity software

Table 2. Number of reads, in the four RNA libraries, mapping the positive and negative orientation of virus segments and control host genes (underlined).

		1-ENDO			2-ENDO			XEO			ORM		
Contig	Length	All	Negative	Positive	All	Negative	Positive	All	Negative	Positive	All	Negative	Positive
***O. maius* Actin**	1743	2338	**7**	**1162**	NA	NA	NA	NA	NA	NA	NA	NA	NA
***O. maius* Tubulin**	2526	21434	**151**	**10566**	NA	NA	NA	NA	NA	NA	NA	NA	NA
***O. maius* ATPase**	3691	48570	**55**	**24230**	NA	NA	NA	NA	NA	NA	NA	NA	NA
***Tulasnella* Actin**	1409	NA	NA	NA	NA	NA	NA	3506	**0**	**1753**	NA	NA	NA
DN37559_c1_g1_i3[a]	2455	203672	**91605**	**10231**	0	0	0	340738	**154367**	**16002**	NA	NA	NA
DN43802_c0_g4_i1[a]	2481	347830	**159583**	**14332**	2	1	0	620664	**285480**	**24852**	NA	NA	NA
DN47822_c2_g3_i2[a]	2087	642822	**473**	**320938**	783912	**264**	**391692**	2	0	1	NA	NA	NA
TuBlV1[b]	12157	NA	NA	NA	NA	NA	NA	NA	NA	NA	13972	**3734**	**3252**
TuBlV1-Nc[b]	780	NA	NA	NA	NA	NA	NA	NA	NA	NA	938	**46**	**423**
TuBlV1-ORF3[b]	3410	NA	NA	NA	NA	NA	NA	NA	NA	NA	4130	**1566**	**499**
TuBlV1-RdRP[b]	7523	NA	NA	NA	NA	NA	NA	NA	NA	NA	7092	**1738**	**1808**

[a] DN37559_c1_g1_i3=Oidiodendron maius splipalmivirus 1 RNA1; DN43802_c0_g4_i1=Oidiodendron maius splipalmivirus 1 RNA2; DN47822_c2_g3_i2=Oidiodendron maius ourmia-like virus contig 1

[b] TuBlV1=Tulasnella bunyavirales-like virus 1. Nc=nucleocapsid; RdRP=RNA-dependent RNA polymerase

4.9 References

Andika I. B., Wei S., Cao C. M., Salaipeth L., Kondo H., Sun L. Y. (2017) Phytopathogenic fungus hosts a plant virus: A naturally occurring cross-kingdom viral infection. *Proceedings of the National Academy of Sciences of the United States of America,* **114**, 12267-12272.

Bai X., Debaud J. C., Schründer J., Meinhardt F. (1997) The ectomycorrhizal basidiomycete Hebeloma circinans harbors a linear plasmid encoding aDNA-and RNA polymerase. *The Journal of general and applied microbiology,* **43**, 273-279.

Bao X., Roossinck M. J. (2013) Multiplexed Interactions: Viruses of Endophytic Fungi. *Advances in Virus Research, Vol 86: Mycoviruses,* **86**, 37-58.

Bertran A., Ciuffo M., Margaria P., Rosa C., Resende R. O., Turina M. (2016) Host-specific accumulation and temperature effects on the generation of dimeric viral RNA species derived from the S-RNA of members of the Tospovirus genus. *Journal of General Virology,* **97**, 3051-3062.

Bian R. L., Andika I. B., Pang T. X., Lian Z. Q., Wei S., Niu E. B., Wu Y. F., Kondo H., Liu X. L., Sun L. Y. (2020) Facilitative and synergistic interactions between fungal and plant viruses. *Proceedings of the National Academy of Sciences of the United States of America,* **117**, 3779-3788.

Blattny C., Kralik O. (1968) A virus disease of Laccaria laccata (Scop. ex Fr.) Cooke and some other fungi. *Ceska Mykologie,* **22**, 161-166.

Bolger A. M., Lohse M., Usadel B. (2014) Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics,* **30**, 2114-2120.

Bonfante P., Genre A. (2010) Mechanisms underlying beneficial plant-fungus interactions in mycorrhizal symbiosis. *Nature Communications,* **1**.

Bradley R., Burt A. J., Read D. J. (1982) The biology of mycorrhiza in the Ericaceae: VIII. The role of mycorrhizal infection in heavy metal resistance. *New Phytologist,* **91**, 197-209.

Chandler J. A., Liu R. M., Bennett S. N. (2015) RNA shotgun metagenomic sequencing of northern California (USA) mosquitoes uncovers viruses, bacteria, and fungi. *Frontiers in Microbiology,* **6**.

Charon J., Grigg M. J., Eden J.-S., Piera K. A., Rana H., William T., Rose K., Davenport M. P., Anstey N. M., Holmes E. C. (2019) Novel RNA viruses associated with Plasmodium vivax in human malaria and Leucocytozoon parasites in avian disease. *Plos Pathogens,* **15**.

Chiapello M., Rodriguez-Romero J., Nerva L., Forgia M., Chitarra W., Ayllon M. A., Turina M. (2020) Putative new plant viruses associated with Plasmopara viticola-infected grapevine samples. *Annals of Applied Biology,* **176**, 180-191.

Couture M., Fortin J. A., Dalpe Y. (1983) Oidiodendron griseum Robak: an endphyte of ericoid mycorrhiza in vaccinum spp. *New Phytologist,* **95**, 375-380.

Daghino S., Di Vietro L., Petiti L., Martino E., Dallabona C., Lodi T., Perotto S. (2019) Yeast expression of mammalian Onzin and fungal FCR1 suggests ancestral functions of PLAC8 proteins in mitochondrial metabolism and DNA repair. *Scientific Reports,* **9**.

Dalpe Y. (1986) Axenic synthesis of ericoid mycorrhiza in *Vaccinium angustigolium* ait. by Oidiodendron species. *New Phytologist,* **103**, 391-396.

Deakin G., Dobbs E., Bennett J. M., Jones I. M., Grogan H. M., Burton K. S. (2017) Multiple viral infections in Agaricus bisporus - Characterisation of 18 unique RNA viruses and 8 ORFans identified by deep sequencing. *Scientific Reports,* **7**.

Dearnaley J. D., Martos F., Selosse M.-A. (2012) 12 orchid mycorrhizas: molecular ecology, physiology, evolution and conservation aspects. *In: Fungal associations*, pp. 207-230. Springer.

DeRisi J. L., Huber G., Kistler A., Retallack H., Wilkinson M., Yllanes D. (2019) An exploration of ambigrammatic sequences in narnaviruses. *Scientific Reports,* **9**.

Dieleman-Van Zaayen A., Igesz O., Finch J. T. (1970) Intracellular appearance and some morphological features of virus-like particles in an ascomycete fungus. *Virology,* **42**, 534-537.

Diep Thi H., Chernomor O., von Haeseler A., Minh B. Q., Le Sy V. (2018) UFBoot2: Improving the Ultrafast Bootstrap Approximation. *Molecular Biology and Evolution,* **35**, 518-522.

Dinan A. M., Lukhovitskaya N. I., Olendraite I., Firth A. E. (2020) A case for a negative-strand coding sequence in a group of positive-sense RNA viruses. *Virus evolution,* **6**, veaa007-veaa007.

Donaire L., Ayllon M. A. (2017) Deep sequencing of mycovirus-derived small RNAs from Botrytis species. *Molecular Plant Pathology,* **18**, 1127-1137.

Espino-Vazquez A. N., Bermudez-Barrientos J. R., Cabrera-Rangel J. F., Cordova-Lopez G., Cardoso-Martinez F., Martinez-Vazquez A., Camarena-Pozos D. A., Mondo S. J., Pawlowska T. E., Abreu-Goodger C., Partida-Martinez L. P. (2020) Narnaviruses: novel players in fungal-bacterial symbioses. *Isme Journal*.

Ferriol I., Vallino M., Ciuffo M., Nigg J. C., Zamora-Macorra E. J., Falk B. W., Turina M. (2018) The Torradovirus-specific RNA2-ORF1 protein is necessary for plant systemic infection. *Molecular Plant Pathology,* **19**, 1319-1331.

Gange A. C., Koricheva J., Currie A. F., Jaber L. R., Vidal S. (2019) Meta-analysis of the role of entomopathogenic and unspecialized fungal endophytes as plant bodyguards. *New Phytologist,* **223**, 2002-2010.

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

Girlanda M., Segreto R., Cafasso D., Liebel H. T., Rodda M., Ercole E., Cozzolino S., Gebauer G., Perotto S. (2011) Photosynthetic Mediterranean meadow orchids feature partial mycoheterotrophy and specific mycorrhizal associations. *American Journal of Botany,* **98**, 1148-1163.

Goertz G. P., Miesen P., Overheul G. J., van Rij R. P., van Oers M. M., Pijlman G. P. (2019) Mosquito Small RNA Responses to West Nile and Insect-Specific Virus Infections in Aedes and Culex Mosquito Cells. *Viruses-Basel,* **11**.

Gorbalenya A. E., Pringle F. M., Zeddam J. L., Luke B. T., Cameron C. E., Kalmakoff J., Hanzlik T. N., Gordon K. H. J., Ward V. K. (2002) The palm subdomain-based active site is internally permuted in viral RNA-dependent RNA polymerases of an ancient lineage. *Journal of Molecular Biology,* **324**, 47-62.

Grabherr M. G., Haas B. J., Yassour M., Levin J. Z., Thompson D. A., Amit I., Adiconis X., Fan L., Raychowdhury R., Zeng Q. D., Chen Z. H., Mauceli E., Hacohen N., Gnirke A., Rhind N., di Palma F., Birren B. W., Nusbaum C., Lindblad-Toh K., Friedman N., Regev A. (2011) Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nature Biotechnology,* **29**, 644-U130.

Grybchuk D., Akopyants N. S., Kostygov A. Y., Konovalovas A., Lye L. F., Dobson D. E., Zangger H., Fasel N., Butenko A., Frolov A. O., Votypka J., d'Avila-Levy C. M., Kulich P., Moravcova J., Plevka P., Rogozin I. B., Serva S., Lukes J., Beverley S. M., Yurchenko V. (2018) Viral discovery and diversity in trypanosomatid protozoa with a focus on relatives of the human parasite Leishmania. *Proceedings of the National Academy of Sciences of the United States of America,* **115**, E506-E515.

Herrero N., Marquez S. S., Zabalgogeazcoa I. (2009) Mycoviruses are common among different species of endophytic fungi of grasses. *Archives of Virology,* **154**, 327-330.

Hillman B. I., Cai G. (2013) The Family Narnaviridae: Simplest of RNA Viruses. *Advances in Virus Research, Vol 86: Mycoviruses,* **86**, 149-176.

Hollings M. (1962) Viruses associated with a die-back disease of cultivated mushroom. *Nature,* **196**, 962-965.

Holmes E. C., Duchene S. (2019) Can Sequence Phylogenies Safely Infer the Origin of the Global Virome? *Mbio,* **10**.

Hrabakova L., Koloniuk I., Petrzik K. (2017) Phomopsis longicolla RNA virus 1-Novel virus at the edge of myco- and plant viruses. *Virology,* **506**, 14-18.

Huttinga H., Wichers H., Dieleman-Van Zaayen A. (1975) Filamentous and polyhedral virus-like particles in Boletus edulis. *Netherlands Journal of Plant Pathology,* **81**, 102-106.

Jacob-Wilk D., Turina M., Van Alfen N. K. (2006) Mycovirus cryphonectria hypovirus 1 elements cofractionate with trans-golgi network membranes of the fungal host Cryphonectria parasitica. *Journal of Virology,* **80**, 6588-6596.

James J. D., Saunders G. C., Owens S. J. (1998) Isolation and partial characterisation of double-stranded RNA-containing viruses of orchid mycorrhizal fungi. *Springer Lab Manual Series; Mycorrhiza manual*, 425-436.

Kadowaki K., Halvorson H. O. (1971) Appearance of a new species of ribonucleic acid during sporulation in *Saccharomyces cerevisiae*. *Journal of Bacteriology,* **105**.

Kaefer S., Paraskevopoulou S., Zirkel F., Wieseke N., Donath A., Petersen M., Jones T. C., Liu S., Zhou X., Middendorf M., Junglen S., Misof B., Drosten C. (2019) Re-assessing the diversity of negative strand RNA viruses in insects. *Plos Pathogens,* **15**.

Kalyaanamoorthy S., Bui Quang M., Wong T. K. F., von Haeseler A., Jermiin L. S. (2017) ModelFinder: fast model selection for accurate phylogenetic estimates. *Nature Methods,* **14**, 587-+.

Kanhayuwa L., Kotta-Loizou I., Oezkan S., Gunning A. P., Coutts H. A. (2015) A novel mycovirus from Aspergillus fumigatus contains four unique dsRNAs as its genome and is infectious as dsRNA. *Proceedings of the National Academy of Sciences of the United States of America,* **112**, 9100-9105.

Kohler A., Kuo A., Nagy L. G., Morin E., Barry K. W., Buscot F., Canback B., Choi C., Cichocki N., Clum A., Colpaert J., Copeland A., Costa M. D., Dore J., Floudas D., Gay G., Girlanda M., Henrissat B., Herrmann S., Hess J., Hogberg N., Johansson T., Khouja H.-R., LaButti K., Lahrmann U., Levasseur A., Lindquist E. A., Lipzen A., Marmeisse R., Martino E., Murat C., Ngan C. Y., Nehls U., Plett J. M., Pringle A., Ohm R. A., Perotto S., Peter M., Riley R., Rineau F., Ruytinx J., Salamov A., Shah F., Sun H., Tarkka M., Tritt A., Veneault-Fourrey C., Zuccaro A., Tunlid A., Grigoriev I. V., Hibbett D. S., Martin F., Mycorrhizal Genomics Initiative C. (2015) Convergent losses of decay mechanisms and rapid turnover of symbiosis genes in mycorrhizal mutualists. *Nature Genetics,* **47**, 410-U176.

Koonin E. V. (1991) The phylogeny of RNA-dependent RNA polymerases of positive-strand RNA viruses. *Journal of General Virology,* **72**, 2197-2206.

Koonin E. V., Dolja V. V., Krupovic M., Varsani A., Wolf Y. I., Yutin N., Zerbini F. M., Kuhn J. H. (2020) Global Organization and Proposed Megataxonomy of the Virus World. *Microbiology and molecular biology reviews : MMBR,* **84**.

Krupovic M., Dolja V. V., Koonin E. V. The LUCA and its complex virome. *Nature Reviews Microbiology*.

Krupovic M., Dolja V. V., Koonin E. V. (2019) Origin of viruses: primordial replicators recruiting capsids from hosts. *Nature Reviews Microbiology,* **17**, 449-458.

Lacourt I., D'Angelo S., Girlanda M., Turnau K., Bonfante P., Perotto S. (2000) Genetic polymorphism and metal sensitivity of Oidiodendron maius strains isolated from polluted soil. *Annals of Microbiology,* **50**, 157-166.

Langmead B., Salzberg S. L. (2012) Fast gapped-read alignment with Bowtie 2. *Nature Methods,* **9**, 357-U354.

Li C.-X., Shi M., Tian J.-H., Lin X.-D., Kang Y.-J., Chen L.-J., Qin X.-C., Xu J., Holmes E. C., Zhang Y.-Z. (2015) Unprecedented genomic diversity of RNA viruses in arthropods reveals the ancestry of negative-sense RNA viruses. *Elife,* **4**.

Lin Y., Zhou J., Zhou X., Shuai S., Zhou R., An H., Fang S., Zhang S., Deng Q. (2020) A novel narnavirus from the plant-pathogenic fungus Magnaporthe oryzae. *Archives of Virology,* **165**, 1235-1240.

Liu L., Xie J., Cheng J., Fu Y., Li G., Yi X., Jiang D. (2014) Fungal negative-stranded RNA virus that is related to bornaviruses and nyaviruses. *Proceedings of the National Academy of Sciences of the United States of America,* **111**, 12205-12210.

Marquez L. M., Redman R. S., Rodriguez R. J., Roossinck M. J. (2007) A virus in a fungus in a plant: Three-way symbiosis required for thermal tolerance. *Science,* **315**, 513-515.

Martino E., Morin E., Grelet G.-A., Kuo A., Kohler A., Daghino S., Barry K. W., Cichocki N., Clum A., Dockter R. B., Hainaut M., Kuo R. C., LaButti K., Lindahl B. D., Lindquist E. A., Lipzen A., Khouja H.-R., Magnuson J., Murat C., Ohm R. A., Singer S. W., Spatafora J. W., Wang M., Veneault-Fourrey C., Henrissat B., Grigoriev I. V., Martin F. M., Perotto S. (2018) Comparative genomics and transcriptomics depict ericoid mycorrhizal fungi as versatile saprotrophs and plant mutualists. *New Phytologist,* **217**, 1213-1229.

Martino E., Perotto S., Parsons R., Gadd G. M. (2003) Solubilization of insoluble inorganic zinc co mpounds by ericoid mycorrhizal fungi derived from heavy metal polluted sites. *Soil Biology & Biochemistry,* **35**, 133-141.

Martino E., Turnau K., Girlanda M., Bonfante P., Perotto S. (2000) Ericoid mycorrhizal fungi from heavy metal polluted soils: their identification and growth in the presence of zinc ions. *Mycological Research,* **104**, 338-344.

Marzano S.-Y. L., Nelson B. D., Ajayi-Oyetunde O., Bradley C. A., Hughes T. J., Hartman G. L., Eastburn D. M., Domier L. L. (2016) Identification of Diverse Mycoviruses through Metatranscriptomics Characterization of the Viromes of Five Major Fungal Plant Pathogens. *Journal of Virology,* **90**, 6846-6863.

Mascia T., Vucurovic A., Minutillo S. A., Nigro F., Labarile R., Savoia M. A., Palukaitis P., Gallitelli D. (2019) Infection of Colletotrichum acutatum and Phytophthora infestans by taxonomically different plant viruses. *European Journal of Plant Pathology,* **153**, 1001-1017.

Milne I., Bayer M., Stephen G., Cardle L., Marshall D. (2016) Tablet: visualizing next-generation sequence assemblies and mappings. *In: Plant Bioinformatics*, pp. 253-268. Springer.

Mollentze N., Streicker D. G. (2020) Viral zoonotic risk is homogenous among taxonomic orders of mammalian and avian reservoir hosts. *Proceedings of the National Academy of Sciences of the United States of America,* **117**, 9423-9430.

Nerva L., Ciuffo M., Vallino M., Margaria P., Varese G. C., Gnavi G., Turina M. (2016) Multiple approaches for the detection and characterization of viral and plasmid symbionts from a collection of marine fungi. *Virus Research,* **219**, 22-38.

Nerva L., Forgia M., Ciuffo M., Chitarra W., Chiapello M., Vallino M., Varese G. C., Turina M. (2019a) The mycovirome of a fungal collection from the sea cucumber Holothuria polii. *Virus Research,* **273**.

Nerva L., Turina M., Zanzotto A., Gardiman M., Gaiotti F., Gambino G., Chitarra W. (2019b) Isolation, molecular characterization and virome analysis of culturable wood fungal endophytes in esca symptomatic and asymptomatic grapevine plants. *Environmental Microbiology,* **21**, 2886-2904.

Nerva L., Varese G. C., Falk B. W., Turina M. (2017) Mycoviruses of an endophytic fungus can replicate in plant cells: evolutionary implications. *Scientific Reports,* **7**.

Ong J. W. L., Li H., Sivasithamparam K., Dixon K. W., Jones M. G. K., Wylie S. J. (2016) Novel Endorna-like viruses, including three with two open reading frames, challenge the membership criteria and taxonomy of the Endornaviridae. *Virology,* **499**, 203-211.

Ong J. W. L., Li H., Sivasithamparam K., Dixon K. W., Jones M. G. K., Wylie S. J. (2017) The challenges of using high-throughput sequencing to track multiple bipartite mycoviruses of wild orchid-fungus partnerships over consecutive years. *Virology,* **510**, 297-304.

Ong J. W. L., Li H., Sivasithamparam K., Dixon K. W., Jones M. G. K., Wylie S. J. (2018) Novel and divergent viruses associated with Australian orchid-fungus symbioses. *Virus Research,* **244**, 276-283.

Osaki H., Sasaki A., Nomiyama K., Tomioka K. (2016) Multiple virus infection in a single strain of Fusarium poae shown by deep sequencing. *Virus Genes,* **52**, 835-847.

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**, 115-128.

Pearson V., Read D. J. (1973) The biology of mycorrhiza in the Ericaceae: I. The isolation of the endophyte and synthesis of mycorrhizas in aseptic culture. *New Phytologist,* **72**, 371-&.

Perez M., Craven R. C., de la Torre J. C. (2003) The small RING finger protein Z drives arenavirus budding: Implications for antiviral strategies. *Proceedings of the National Academy of Sciences of the United States of America,* **100**, 12978-12983.

Perotto S., ActisPerino E., Perugini J., Bonfante P. (1996) Molecular diversity of fungi from ericoid mycorrhizal roots. *Molecular Ecology,* **5**, 123-131.

Petrzik K., Sarkisova T., Starý J., Koloniuk I., Hrabáková L., Kubešová O. (2016) Molecular characterization of a new monopartite dsRNA mycovirus from mycorrhizal Thelephora terrestris (Ehrh.) and its detection in soil oribatid mites (Acari: Oribatida). *Virology,* **489**, 12-19.

R Development Core Team R. (2011) R: A language and environment for statistical computing. *In*.: R foundation for statistical computing Vienna, Austria.

Rastgou M., Habibi M. K., Izadpanah K., Masenga V., Milne R. G., Wolf Y. I., Koonin E. V., Turina M. (2009) Molecular characterization of the plant virus genus Ourmiavirus and evidence of inter-kingdom reassortment of viral genome segments as its possible route of origin. *Journal of General Virology,* **90**, 2525-2535.

Read D. J. (1974) Pezizella ericae sp. nov., the perfect state of a typical mycorrhizal endophyte of Ericaceae. *Transactions of the British Mycological Society,* **63**, 381-&.

Read D. J., Stribley D. P. (1973) Effect of mycorrhizal infection on nitrogen and phosphorus nutrition of ericaceous plants. *Nature-New Biology,* **244**, 81-82.

Richaud A., Frezal L., Tahan S., Jiang H., Blatter J. A., Zhao G., Kaur T., Wang D., Felix M.-A. (2019) Vertical transmission in Caenorhabditis nematodes of RNA molecules encoding a viral RNA-dependent RNA polymerase. *Proceedings of the National Academy of Sciences of the United States of America,* **116**, 24738-24747.

Rodriguez R. J., White J. F., Jr., Arnold A. E., Redman R. S. (2009) Fungal endophytes: diversity and functional roles. *New Phytologist,* **182**, 314-330.

Rohwer F., Prangishvili D., Lindell D. (2009) Roles of viruses in the environment. *Environmental Microbiology,* **11**, 2771-2774.

Roossinck M. J. (2019) Evolutionary and ecological links between plant and fungal viruses. *New Phytologist,* **221**, 86-92.

Ryan F. (2009) *Virolution*: Collins.

Sato Y., Shamsi W., Jamal A., Bhatti M. F., Kondo H., Suzuki N. (2020) Hadaka Virus 1: a Capsidless Eleven-Segmented Positive-Sense Single-Stranded RNA Virus from a Phytopathogenic Fungus, Fusarium oxysporum. *mBio,* **11**.

Shi M., Lin X.-D., Tian J.-H., Chen L.-J., Chen X., Li C.-X., Qin X.-C., Li J., Cao J.-P., Eden J.-S. (2016) Redefining the invertebrate RNA virosphere. *Nature,* **540**, 539-543.

Sievers F., Wilm A., Dineen D., Gibson T. J., Karplus K., Li W., Lopez R., McWilliam H., Remmert M., Soeding J., Thompson J. D., Higgins D. G. (2011) Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Molecular Systems Biology,* **7**.

Smith S. E., Read D. J. (2008) Mycorrhizal Symbiosis, 3rd Edition. *Mycorrhizal Symbiosis, 3rd Edition*, 1-787.

Steitz T. A. (1998) Structural biology - A mechanism for all polymerases. *Nature,* **391**, 231-232.

Stielow B., Menzel W. (2010) Complete nucleotide sequence of TaV1, a novel totivirus isolated from a black truffle ascocarp (Tuber aestivum Vittad.). *Archives of Virology,* **155**, 2075-2078.

Suttle C. A. (2007) Marine viruses - major players in the global ecosystem. *Nature Reviews Microbiology,* **5**, 801-812.

Takahashi H., Fukuhara T., Kitazawa H., Kormelink R. (2019) Virus Latency and the Impact on Plants. *Frontiers in Microbiology,* **10**.

te Velthuis A. J. W. (2014) Common and unique features of viral RNA-dependent polymerases. *Cellular and Molecular Life Sciences,* **71**, 4403-4420.

Trifinopoulos J., Lam-Tung N., von Haeseler A., Minh B. Q. (2016) W-IQ-TREE: a fast online phylogenetic tool for maximum likelihood analysis. *Nucleic Acids Research,* **44**, W232-W235.

Turina M., Prodi A., Van Alfen N. K. (2003) Role of the Mf1-1 pheromone precursor gene of the filamentous ascomycete Cryphonectria parasitica. *Fungal Genetics and Biology,* **40**, 242-251.

Vainio E. J., Jurvansuu J., Streng J., Rajamaki M. L., Hantula J., Valkonen J. P. T. (2015) Diagnosis and discovery of fungal viruses using deep sequencing of small RNAs. *Journal of General Virology,* **96**, 714-725.

Vainio E. J., Pennanen T., Rajala T., Hantula J. (2017) Occurrence of similar mycoviruses in pathogenic, saprotrophic and mycorrhizal fungi inhabiting the same forest stand. *FEMS microbiology ecology,* **93**.

Vallino M., Zampieri E., Murat C., Girlanda M., Picarella S., Pitet M., Portis E., Martino E., Perotto S. (2011) Specific regions in the Sod1 locus of the ericoid mycorrhizal fungus Oidiodendron maius from metal-enriched soils show a different sequence polymorphism. *Fems Microbiology Ecology,* **75**, 321-331.

Waldron F. M., Stone G. N., Obbard D. J. (2018) Metagenomic sequencing suggests a diversity of RNA interference-like responses to viruses across multicellular eukaryotes. *Plos Genetics,* **14**.

Weiss M., Waller F., Zuccaro A., Selosse M.-A. (2016) Sebacinales - one thousand and one interactions with land plants. *New Phytologist,* **211**, 20-40.

Wesolowski M., Wickner R. B. (1984) Two new double-stranded RNA molecules showing non-mendelian inheritance and heat inducibility in Saccharomyces cerevisiae. *Molecular and Cellular Biology,* **4**, 181-187.

Wolf Y. I., Kazlauskas D., Iranzo J., Lucia-Sanz A., Kuhn J. H., Krupovic M., Dolja V. V., Koonin E. V. (2018) Origins and Evolution of the Global RNA Virome. *Mbio,* **9**.

Wu F., Zhao S., Yu B., Chen Y.-M., Wang W., Song Z.-G., Hu Y., Tao Z.-W., Tian J.-H., Pei Y.-Y., Yuan M.-L., Zhang Y.-L., Dai F.-H., Liu Y., Wang Q.-M., Zheng J.-J., Xu L., Holmes E. C., Zhang Y.-Z. (2020) A new coronavirus associated with human respiratory disease in China. *Nature*.

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. *Proceedings of the National Academy of Sciences of the United States of America,* **107**, 8387-8392.

Zoll J., Verweij P. E., Melchers W. J. G. (2018) Discovery and characterization of novel Aspergillus fumigatus mycoviruses. *Plos One,* **13**.

5 Ongoing experiments: phenotypes evaluation of the viral infections

Our work has shown the occurrence of new mycovirus phylogenetic lineages that could only partially be characterized molecularly: nevertheless, most of these viruses have been described in fungal isolates (and not metagenomic samples) and this will allow further biological characterization and a more complete molecular one, possibly with the necessary intermediate step of establishing an infectious clone. Here below are summarized the main research fields we are currently pursuing. Furthermore, our initially planned work on *Aspergillus ochraceus* mycovirus-regulated mycotoxin production had to be put on hold for lack of genomic data (published genome information on this species, turned out to be a wrong taxonomical assignment). We collaborated to the *A. ochraceous* genome sequence and this will give us the possibility to further pursue this line of research as detailed below.

5.1 Investigation of ambivirus induced phenotypic effects

The discovery and characterization of several ambiviruses in different fungal hosts raised curiosity about the possible biological effects that these could cause to their fungal hosts. Moreover, we also could not exclude that ambiviruses could harbor more genomic segments than the ones we have been able to identify so far, based on homology searches. With the aim of gathering further evidence for the characterization of the complete structure of genomes of ambiviruses, we conducted a new bioinformatic characterization on Cryphonectria parasitica ambivirus 1. Since the transcriptomic samples discussed in chapter 3 were composed by pooled RNAs from several fungal isolates, we performed RNAseq analysis on total RNA from the CpaV1 infected ACP 34. With a transcriptomic sample obtained from a single infected isolate, we compared the obtained transcriptome to the reference from *Cryphonectria parasitica* EP155 strain (Crouch *et al.*, 2020) to search for evidence of other contigs associated to the ambivirus. After a Blastx analysis using as query ACP 34 transcriptome and as a reference the annotated proteins from EP155, we retrieved the contigs without a clear match (all the results with an e value below 0,00001) and we further screen them through a Blastn approach against EP155 genomic sequence (to eliminate possible DNA contamination in ACP 34 sample); then we performed a Blastx analysis against the NCBI nr database, to check for the presence of uncharacterized proteins of fungal origin. The results were filtered to obtain contigs over 500 nucleotides longs: this analysis resulted in 12 ORFAN contigs. We then designed qRT-PCR primers on the twelve contigs and checked their presence in cDNA and DNA from *C. parasitica* ACP34. Results showed that the twelve contigs were found both in the cDNA and the DNA from the tested isolate, suggesting a genomic origin of the contigs that was probably not noticed as we used a masked version of the reference assembly throughout our bioinformatic pipeline. With this premises, we can therefore exclude that these are ambivirus-associated contigs. Even if a complete characterization of CpaV1 sequence should be obtained repeating the same analysis against transcriptomic data obtained from an isogenic non-infected *C. parasitica* isolate, the result support our hypothesis of a monopartite genomic structures for ambiviruses.

The fact that, for most of the ambiviruses detected, we have been able to identify what is probably a dimeric form of the sequence complicates the strategy to produce an infectious clone (that could be easily transfected at least on *Cryphonectria parasitica*, where easy and efficient protoplast transfection protocols are available) as it is hard to understand which is the correct monomeric sequence of the virus. For these reasons, we tried to obtain isogenic virus-infected and virus-free isolates starting from one of the reference *C. parasitica* isolates infected with CpaV1. From *C. parasitica* isolate ACP34, we collected the conidia by filtration and we performed qRT-PCR detection of CpaV1 on single-conidia colonies. We have been able to identify five CpaV1 infected and five non-infected colonies that are currently under investigation to understand if any phenotype change could be observed through direct comparisons.

5.2 Splipalmivirus infectious clone construction and phenotype evaluation in the ericoid mycorrhizal context

For the first time, our study reported a virus where the conserved palm domain appears split between two proteins encoded by distinct genomic segments. As we described in the previous chapters, this evidence is highly relevant not only because, to our knowledge, the palm domain is conserved in all RNA viruses, but also because, for the first time, it shows a new RdRP catalytic mechanism for RNA viruses. Since the high interest related to this virus named Oidiodendron maius splipalmivirus 1 (OmSPV1) and to his host *Oidiodendron maius* (an ericoid mycorrhizal fungus), we started the construction of an infectious clone to initiate the replication in a reference isolate from infectious run-off transcripts. The aim of this work will be to demonstrate that the two viral fragments are required together to infect the host (thus demonstrating that replication occurs when all the conserved palm motifs are present); moreover, an efficient infectious clone could allow the comprehension of virus-driven phenotypes eventually present in the context of the association between *O. maius* and ericoid plants. The strategy pursued to build the infectious clone is similar to the one described for other viruses from the Phylum *Lenarviricota* (Esteban *et al.*, 2005; Wang *et al.*, 2020): the full-length sequences from OmSPV1 RNA1 and RNA2 were amplified and cloned inside a pCR blunt vector from Invitrogen. Using the obtained vectors as template, a PCR reaction was set up to produce the template for RNA transcription using a forward primer containing a T7 promoter upstream of the 5' end of the virus segment. The obtained PCR product was then used to produce the viral RNAs through T7 transcription. Viral RNAs were transfected in *O. maius* protoplasts and obtained fungal colonies are currently under testing through qRT-PCR to detect positively infected isolates. If the preliminary experiments will show interesting results, allowing us to obtain isogenic isolates virus-infected and virus-free, we will then characterize the fungal phenotypes both during its saprophytic lifestyle and during mycorrhization.

5.3 Vivivirus characterization through comparisons of conserved terminal sequences

In the context of other studies conducted in our laboratory, a new mycoviral lineage was characterized and viruses belonging to this group were named viviviruses (Chiapello *et al.*, 2020). These viruses are characterized by two genomic fragments carrying one ORF each: one of the ORF encodes for a protein showing the RdRp domain and a methyltransferase domain while the ORF on the second fragment encodes for a protein with a conserved helicase domain. Homologues to the two RNAs from viviviruses were already discovered and deposited as Luckshill virus (NCBI ID: AWA82251) and Cyril virus (NCBI ID: AWA82277), but they were not reported as part of the same genome (Medd *et al.*, 2018). Indeed, Chiapello and coworkers were the first to observe through 3'-5' ends comparison and qRT-PCR data that the two fragments are together part of the same viral genome. In the last months, many submissions on the NCBI database brought to an increase in the number of viviviruses discovered and through bioinformatic pipelines another fragment associated to viviviruses has been reported related to the virus named Aspergillus fumigatus RNA virus 1 (AfuRV1) (Chiba *et al.*, 2020). Recently we were able to identify another vivivirus infecting *Aspergillus flavus* obtaining for the first-time data about the vivivirus viral particles morphology through purification on sucrose cushion. The infected isolate (called MN) was investigated for viral presence using the same approach explained in chapter 2 to 4 and was found harboring two different viruses: a tymo-flexy virus and a vivivirus. Transmission electronic microscopy analysis performed on fungal samples purified as shown in previous work (Nerva *et al.*, 2016) showed two kind of particles: a long-flexuous type compatible with tymo-flexy like mycoviruses presence (data not shown) and an isometric type that we associated to the vivivirus that we isolated through step sucrose gradient and precipitation with ultracentrifugation (Fig1C). The evidence of a particle associated to vivivirus infection let us focus on the identification of the putative capsid protein; since the genomic fragments share a conserved motif, we first searched for other contigs from the transcriptomic sample used to identify the virus by looking at conserved nucleotide motifs on contig's ends through blastn. With this analysis we identified eight new contigs (Fig1A-B), each of them encoding for a putative protein with unknown function; we then tested the contigs through qRT-PCR to check if their presence is strictly correlated to the two segments already characterized for the viral genome, and indeed, we confirmed their presence only in the cDNAs obtained from the infected isolate MN and not in MN's DNA or in cDNA from non-infected *A. flavus*. Blastp analysis between hypothetical protein from each of the identified contigs gave no hit as result against the NCBI nr database (setting an e value of 0.00001) but partial conservation was observed for two proteins comparing them to AfuRV1 RNA3 hypothetical protein (Table 1). The same approach was then performed to search for vivivirus-related contigs in metagenomics samples from Chiapello's work, obtaining also in this analysis eight contigs encoding for one protein each. Blast comparison between *A. flavus* vivivirus and *P. viticola* related vivivirus showed partial conservation only for two of the eight proteins

(Table 2), suggesting poor conservation between segments encoded by different vivivirus. In conclusion, our data present a vivivirus from *A. flavus* with at least 10 genome segments proving that vivivirus could show a genomic organization that is way more complicated than initially characterized based on blast similarity searches. However, we still don't have a direct evidence of the segment encoding the putative viral coat protein, that could be one of the segments we have already identified, or a protein from a still unidentified genome segment. To answer this question, we are now purifying the viral particles to perform protein gel electrophoresis in denaturing condition to separate the specific band related to the vivivirus coat protein. We will perform mass spectrometry analysis to associate the protein band to its sequence retrieving the contig of interest and reporting for the first-time information about morphology and genome organization of viviviruses.

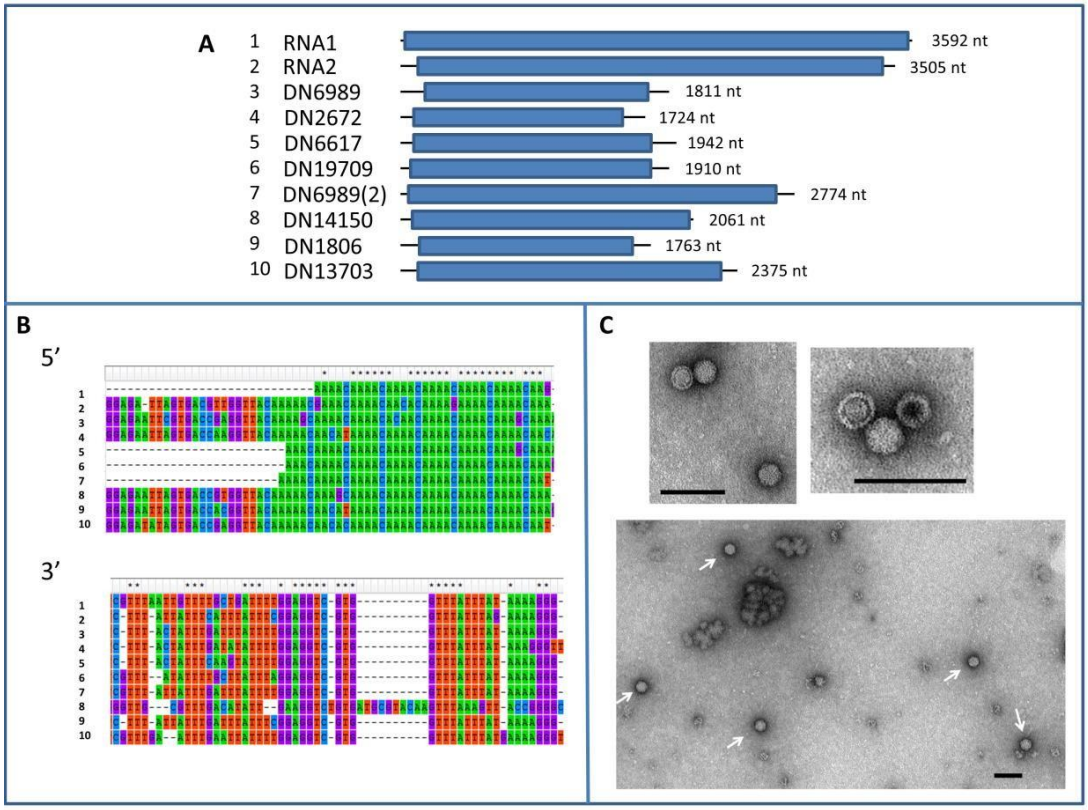


Figure 1. A) Schematic representation of the genomic organization of Aspergillus flavus vivivirus, the full-length sequence is represented by the black line and the blue boxes are the ORFs, contigs sequence can be obtained from the following URL: (https://drive.google.com/drive/folders/1CxPWI3TZsiQV_2W7wiqRjdGGCDdmI-Ar?usp=sharing). **B)** 5' and 3' ends alignment showing conservation between sequences. **C)** Purified viral particles from Aspergillus flavus vivivirus, the scale bar is 100 nm long, the viral particles have a diameter of around 35 nm.

Table 1. Tblastn search using as query the RNA3 hypothetical protein from Aspergillus flavus RNA virus 1 against all the genomic fragment from Aspergillus flavus vivivirus.

Query	*A. flavus* MN	Max Score	Total Score	Query Cover	E value	Per. Ident	Acc. Len
AfrV1 RNA3	DN19709	130	166	66%	3E-35	38.67%	1910
AfrV1 RNA3	DN6617	102	172	56%	6E-26	35.50%	1942

Table 2. Blastx search using as query the eight contigs from *A. flavus* vivivirus against the proteins encoded by the eight contigs identified in *P. viticola*-related vivivirus.

Query	PV Vivivirus	Max Score	Total Score	Query Cover	E value	Per. Ident	Acc. Len
DN14150	contig8	222	222	62%	2,00E-65	34.70%	830
DN13703	contig12	63.2	63.2	51%	3,00E-13	25.80%	752

5.4 Molecular investigation of the mycoviral induced ochratoxin A overproduction in *Aspergillus ochraceus*

An interesting example of mycovirus-induced biological effects on its fungal host was presented by our group some years ago after identifying an *Aspergillus ochraceus* isolate infected with the partitivirus Aspergillus ochraceus virus (AoV) (Nerva *et al.*, 2019b). *A. ochraceus* is known to be a potential mycotoxin producer, contaminating food with the secondary metabolites ochratoxin A (OTA). Our studies brought us to investigate if the mycoviral infection could change the OTA production somehow, generating isogenic fungal isolates virus-infected and virus-free through purified viral particles transfection in fungal protoplasts. After a comparison of the OTA production in the two condition we could observe higher OTA concentration in the AoV infected isolates. This phenotype was specific to AoV infection; indeed, the same overproduction was not observed when quantifying OTA in *A. ochraceus* infected with a different mycovirus called Penicillium aurantiogriseum totivirus 1 (Nerva *et al.*, 2019a). This result rose our interest on the molecular mechanisms exploited by the virus to modify the host metabolism, so we proceed in the characterization of the trascriptomic expression of the AoV-infected isolate to identify mycovirus modulated genes. Our hypothesis was that the regulation could occur through viral sRNAs produced by the DICER-AGO pathway through fungal antiviral response; these viral sRNAs could then target endogenous genes to modulate their expression. The evaluation of our theory was complicated by the poor characterization of the OTA biosynthetic pathway in *A. ochraceus*: the genes responsible for the mycotoxin production are well known for many fungi from the *Aspergillus* genus but not for *A. ochraceus*. Thus, we started a collaboration with ISPA-CNR institute and the JGI institute for de novo sequencing of our reference *A. ochraceus* genome to perform a complete annotation and identify the cluster responsible for OTA biosynthesis. At the same time, we proceeded with the RNA sequencing and sRNA sequencing from both AoV and PaTV1 infected *A. ochraceus* and we also performed degradome analysis (German *et al.*, 2009) on our samples to evaluate evidence of the action of viral sRNAs on potential fungal targets. At this moment, our partner at the JGI completed the sequencing and assembly of our *A. ochraceus* isolate genome, and its annotation is currently being performed. With our RNAseq data, we generated some preliminary de novo transcriptome allowing searching for OTA biosynthetic genes comparing them to characterized genes from other *Aspergillus* species (Gallo *et al.*, 2017). As soon as the genome annotation will be completed, we will perform differential expression analysis on our RNAseq data and we will analyze them together with viral sRNA data and degradome to search for any possible target allowing us to hypothesize the molecular mechanisms that allow AoV to modify metabolites production in the host.

5.5 Acknowledgements

The works presented in this chapter are related to different collaborations. Aspergillus flavus vivivirus was identified and characterized through a collaboration with the University of Parma (Dr. Francesca Degola) and the Research Centre for Viticulture and Enology, Council for Agricultural Research and Economics, Conegliano, Italy (Dr. Luca Nerva). The molecular characterization of AoV infected *A. ochraceus* is realized with a collaboration between the Universtity of Turin – MUT (Prof. Giovanna Cristina Varese), the PNNL (Prof. Scott E. Baker) and the ISPA-CNR institute (Dr. Giancarlo Perrone).

5.6 References

Chiapello M., Rodriguez-Romero J., Ayllon M. A., Turina M. (2020) Analysis of the virome associated to grapevine downy mildew lesions reveals new mycovirus lineages. *Virus evolution,* **6**, veaa058-veaa058.

Chiba Y., Yaguchi T., Urayama S.-i., Hagiwara D. (2020) Discovery of divided RdRp sequences and a hitherto unknown genomic complexity in fungal viruses. *In*. BioRxiv: Cold Spring Harbor Laboratory.

Crouch J. A., Dawe A., Aerts A., Barry K., Churchill A. C. L., Grimwood J., Hillman B., Milgroom M. G., Pangilinan J., Smith M., Salamov A., Schmutz J., Yadav J., Grigoriev I. V., Nuss D. (2020) Genome sequence of the chestnut blight fungus Cryphonectria parasitica EP155: A fundamental resource for an archetypical invasive plant pathogen. *Phytopathology,* **110**, 6.

Esteban R., Vega L., Fujimura T. (2005) Launching of the yeast 20 S RNA narnavirus by expressing the genomic or antigenomic viral RNA in vivo. *Journal of Biological Chemistry,* **280**, 33725-33734.

Gallo A., Ferrara M., Perrone G. (2017) Recent advances on the molecular aspects of ochratoxin A biosynthesis. *Current Opinion in Food Science,* **17**, 49-56.

German M. A., Luo S., Schroth G., Meyers B. C., Green P. J. (2009) Construction of Parallel Analysis of RNA Ends (PARE) libraries for the study of cleaved miRNA targets and the RNA degradome. *Nature Protocols,* **4**, 356-362.

Medd N. C., Fellous S., Waldron F. M., Xuereb A., Nakai M., Cross J. V., Obbard D. J. (2018) The virome of Drosophila suzukii, an invasive pest of soft fruit. *Virus Evolution,* **4**.

Nerva L., Chitarra W., Siciliano I., Gaiotti F., Ciuffo M., Forgia M., Varese G. C., Turina M. (2019a) Mycoviruses mediate mycotoxin regulation in Aspergillus ochraceus. *Environmental Microbiology,* **21**, 1957-1968.

Nerva L., Ciuffo M., Vallino M., Margaria P., Varese G. C., Gnavi G., Turina M. (2016) Multiple approaches for the detection and characterization of viral and plasmid symbionts from a collection of marine fungi. *Virus Research,* **219**, 22-38.

Nerva L., Forgia M., Ciuffo M., Chitarra W., Chiapello M., Vallino M., Varese G. C., Turina M. (2019b) The mycovirome of a fungal collection from the sea cucumber Holothuria polii. *Virus research,* **273**, 197737-197737.

Wang Q., Mu F., Xie J., Cheng J., Fu Y., Jiang D. (2020) A Single ssRNA Segment Encoding RdRp Is Sufficient for Replication, Infection, and Transmission of Ourmia-Like Virus in Fungi. *Frontiers in Microbiology,* **11**, 379.

6 General conclusions

The studies presented in this work gave us the opportunity to further investigate inside the dark matter constituted by ORFan contigs assembled from ribosome depleted total RNA, revealing new viral sequences, or new virus types altogether; this approach widened the RNA virus world to include some examples of viral entities that would not be detected with normal homology-based methods. Indeed, after just few months from the publication of the results presented here, other authors reported new members of the proposed ambivirus group infecting *Armillaria spp.* fungi (Linnakoski *et al.*, 2021) and sequences of RNA2 from narna-like splipalmiviruses (Chiba *et al.*, 2021) demonstrating that these viral taxa are probably well spread throughout the fungal kingdom. Again, many of these interesting results on new viral taxa were allowed by the possibility to compare different metatranscriptomic samples and the by possibility to perform comparisons among different virus-infected fungi, showing how our comprehension of the viral world is still partial and how bioinformatic data could be useful to assign new virus types to the virosphere.

Nevertheless, the discovery of new viral taxa isn't the only remarkable result obtained through this work. By characterizing new members of known taxa, a better comprehension of the evolutionary implications is obtained for mycoviruses and for the viral world in general. Indeed, a recombination event has just been speculated by Abdoulaye and coworkers (Abdoulaye *et al.*, 2021) when analyzing a new hypovirus infecting *R. solani* showing two different helicase domains: a first domain typical of hypoviruses and a second related to *Nidovirales* helicases. Authors hypothesize that the *Nidovirales* related helicase could originate from a recombination between an hypovirus and a Rhizoctonia solani putative virus 2-like sequence (described here for the first time in chapter 2) finding a possible evolutionary explanation to a previously undescribed genome organization.

The close spatial association between plant and fungi, with numerous exchanges that include small RNA as a possible inter-kingdom messengers (Wang *et al.*, 2017), prompted a number of studies that looked at possible example of experimental confirmation of inter-kingdom horizontal virus transfers (Nerva *et al.*, 2017). These works proved that plant viruses could infect fungi , and that fungi could also become biological vector of plant viruses, if co-infected with a mycovirus that facilitates plant virus accumulation (Andika *et al.*, 2017); moreover, synergistic interactions could happen between plant and fungal viruses, both when infecting the plant and the fungus (Bian *et al.*, 2020). This work rose further interest on the close interaction occurring between plants and synergistic or biotrophic fungi (such as mycorrhizal fungi or obligate biotrophic fungal/oomycetes pathogen, such as agents of powdery and downy mildews) that are characterized by a constant exchange of resources between the two organisms, which could represent a crucial point in the emergence of viral diversity thanks to the continuity created between the host's viromes. Taken together these results, it appears that the study of the virome

associated to fungi is useful for understanding virus phylogenetic relationship, and virus origin in particular, since mycovirus are likely at the root of some evolutionary trajectories that allowed emergence of some plant viruses. Examples of these events are represented by the identification of a plant specific clade inside the family *Mitoviridae* (Nerva *et al.*, 2019) or the close relatedness between fungal and plant viruses from the families *Partitiviridae* and *Endornaviridae* (Roossinck *et al.*, 2011; Nibert *et al.*, 2014).

In this respect much attention was recently paid to virus spillover among species, resulting in destructive epidemic and pandemics such as that caused by Ebolavirus or by SarsCov2: our work, and the work of others looking at viromes of different fungi in different environments has greatly increased the density of the sampling of mycoviruses even though the biodiversity of mycoviruses is still far from being saturated. These studies are pivotal in revealing evidence of spillover events in fungi, giving us a theoretical framework for predicting such events. This could be important also in the context of the risk assessment of a biocontrol strategy based on mycoviruses; the successful experience of biocontrol of the chestnut blight disease in Europe through Cryphonectria hypovirus 1 diffusion increased the interest on the identification and dissemination of possible mycovirus biocontrol agents, and the lack of extracellular phase brought to the idea that mycoviruses could only be spread between compatible mycelia. Nevertheless, few evidence are available on the possibility to detect the same virus infecting unrelated fungi (Deng *et al.*, 2003; Khalifa, 2019), therefore showing evidence of natural occurrence of mycoviruses crossing inter-species barriers; this possibility will further be evaluated as soon as more biological studies associating a mycovirus with a specific fungal host are available.

Lastly, even If still preliminary, our focus on the biological function of the identified viruses will hopefully help in understanding the often-cryptic viral role in the environment and in interactions among organisms. As mentioned in chapter 1, many examples of mycovirus-induced phenotypes have been reported; in this last year, the association between the endophytic lifestyle and mycoviral infection was shown twice for different fungi (*Sclerotinia sclerotiorum* and *Pestalotiopsis theae*) infected with unrelated mycoviruses (a ssDNA virus and a crysovirus) (Zhang *et al.*, 2020; Tian *et al.*, 2020; Zhou *et al.*, 2021). In both cases, the infected fungi showed a pathogenic lifestyle when the isolates were cured from the infection. The molecular aspects of these interactions are still not fully clear, but the promising results showed so far increased our interest on the possible mutualistic biological role of the viruses we identified giving us the opportunity to further investigate this topic through many different viral and fungal model, both pathogens or symbionts.

6.1 References

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

Andika I. B., Wei S., Cao C. M., Salaipeth L., Kondo H., Sun L. Y. (2017) Phytopathogenic fungus hosts a plant virus: A naturally occurring cross-kingdom viral infection. *Proceedings of the National Academy of Sciences of the United States of America,* **114**, 12267-12272.

Bian R. L., Andika I. B., Pang T. X., Lian Z. Q., Wei S., Niu E. B., Wu Y. F., Kondo H., Liu X. L., Sun L. Y. (2020) Facilitative and synergistic interactions between fungal and plant viruses. *Proceedings of the National Academy of Sciences of the United States of America,* **117**, 3779-3788.

Chiba Y., Oiki S., Yaguchi T., Urayama S., Hagiwara D. (2021) Discovery of divided RdRp sequences and a hitherto unknown genomic complexity in fungal viruses. *Virus Evolution,* **7**.

Deng F., Xu R., Boland G. J. (2003) Hypovirulence-associated double-stranded RNA from Sclerotinia homoeocarpa is conspecific with Ophiostoma novo-ulmi mitovirus 3a-Ld. *Phytopathology,* **93**, 1407-1414.

Khalifa M., E., MacDiarmid, Robin, M. (2019) A Novel Totivirus Naturally Occurring in Two Different Fungal Genera. *Frontiers in Microbiology*.

Linnakoski R., Sutela S., Coetzee M. P. A., Duong T. A., Pavlov I. N., Litovka Y. A., Hantula J., Wingfield B. D., Vainio E. J. (2021) Armillaria root rot fungi host single-stranded RNA viruses. *Scientific Reports,* **11**, 1-15.

Nerva L., Varese G. C., Falk B. W., Turina M. (2017) Mycoviruses of an endophytic fungus can replicate in plant cells: evolutionary implications. *Scientific reports,* **7**, 1908-1908.

Nerva L., Vigani G., Di Silvestro 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**.

Nibert M. L., Ghabrial S. A., Maiss E., Lesker T., Vainio E. J., Jiang D., Suzuki N. (2014) Taxonomic reorganization of family Partitiviridae and other recent progress in partitivirus research. *Virus Research,* **188**, 128-141.

Roossinck M. J., Sabanadzovic S., Okada R., Valverde R. A. (2011) The remarkable evolutionary history of endornaviruses. *Journal of general virology,* **92**, 2674-2678.

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. *Isme Journal*.

Wang M., Weiberg A., Dellota E., Yamane D., Jin H. L. (2017) Botrytis small RNA Bc-siR37 suppresses plant defense genes by cross-kingdom RNAi. *Rna Biology,* **14**, 421-428.

Zhang H., Xie J., Fu Y., Cheng J., Qu Z., Zhao Z., Cheng S., Chen T., Li B., Wang Q., Liu X., Tian B., Collinge D. B., Jiang D. (2020) A 2-kb Mycovirus Converts a Pathogenic Fungus into a Beneficial Endophyte for Brassica Protection and Yield Enhancement. *Molecular Plant,* **13**, 1420-1433.

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*.