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

Massimo Turina[1], Stefano Ghignone[1], Nausicaa Astolfi[2], Alessandro Silvestri[2], Paola Bonfante[2] and Luisa Lanfranco[2]*

[1] Institute for Sustainable Plant Protection – CNR Torino, Strada delle Cacce 73, 10131 Torino (Italy)

[2] Department of Life Sciences and Systems Biology – University of Torino, Viale P.A. Mattioli 25, 10125 Torino (Italy)

***Corresponding Author**

Luisa Lanfranco

Department of Life Sciences and Systems Biology – University of Torino, Viale P.A. Mattioli 25, 10125 Torino (Italy)

e-mail: luisa.lanfranco@unito.it

Running title: The virome of *Gigaspora margarita*

Originality-Significance statement: this paper reports the first virome characterization of *Gigaspora margarita*, an arbuscular mycorrhizal fungus already shown to harbor a complex associated bacterial microbiota therefore adding a further layer of complexity to this organism; furthermore, it reports for the first time the occurrence of cDNA fragments corresponding to replicating non retroviral RNA viruses outside the phylum arthropoda, a discovery that, in those organisms, has uncovered new anti-viral defence pathways.

SUMMARY

Arbuscular Mycorrhizal Fungi (AMF) are key components of the plant microbiota. AMF genetic complexity is increased by the presence of endobacteria which live inside many species. A further component of such complexity is the virome associated to AMF, whose knowledge is still very limited. Here, by exploiting transcriptomic data we describe the virome of *Gigaspora margarita.* A BLAST search for viral RNA-dependent RNA polymerases sequences allowed the identification of four mitoviruses, one Ourmia-like narnavirus, one Giardia-like virus, and two sequences related to Fusarium mycoviruses. Northern blot and RT-PCR confirmed the authenticity of all the sequences with the exception of the Fusarium-related ones. All the mitoviruses are replicative and functional since both positive strand and negative strand RNA are present. The abundance of the viral RNA molecules was not regulated by the presence or absence of *Candidatus* Glomeribacter gigasporarum, the endobacterium hosted by *G. margarita,* with the exception of the Ourmia-like sequence which is absent in bacteria-cured spores. In addition, we report, for the first time, DNA fragments corresponding to mitovirus sequences associated to the presence of viral RNA. These sequences are integrated in neither the mitochondrial nor the nuclear genome, but likely exist as extrachromosomal fragments.

INTRODUCTION

The majority of land plants, including many crops and horticultural species, establish root symbiotic interactions with a small group of soil fungi, the so called Arbuscular Mycorrhizal Fungi (AMF), which belong to the Glomeromycotina (Spatafora *et al.,* 2016). As very ancient and widespread in nature, AMF are key components of the plant microbiota. The benefits that AMF provide to host plants, which include an improved mineral nutrition and an increased tolerance to biotic and abiotic stresses, have raised the interest towards their exploitation as key components of sustainable low input agricultural practices (Berruti *et al.,* 2016; Rodriguez and Sanders, 2015).

AMF obligate biotrophic nature and the lack of stable genetic transformation protocols have made them a recalcitrant biological system to study. The recent publication of genomic data for *Rhizophagus irregularis* (Tisserant *et al.,* 2013; Liu *et al.,* 2014; Ropars *et al.,* 2016) and gene repertoires of other species such as R. *clarus* (Sędzielewska Toro and Brachman, 2016), *Gigaspora margarita* (Salvioli *et al.,* 2016) and *Gigaspora rosea* (Tang *et al.,* 2016) have shed light on the biology and evolution of AMF (Kamel *et al.,* 2016). These findings have allowed to propose a new phylogenetic classification (Spatafora *et al.,* 2016) and to infer the occurrence of a sexual cycle in AMF (Corradi and Brachmann, 2017), for long considered clonal organisms. *R. irregularis* possesses a large genome, compared to other fungi, with about 28,000 protein encoding genes and rich in transposable elements (Tisserant *et al.,* 2013; Liu *et al.,* 2014). One striking feature is the almost complete lack of glycoside hydrolases, a sign of the inability of the fungus to degrade plant cell wall polysaccharides, possibly as a strategy to guarantee an intimate and compatible interaction. The genome data have also been instrumental to describe a large repertoire of putative secreted proteins to look for molecular tools exploited by the fungus to colonize the host plant (Kamel *et al.,* 2017). Very recently, new findings have shed light on their nature of these obligate biotrophs. The lack of a fatty acid synthase gene in the genome of *R. irregularis* is mirrored by fatty acid auxotrophy: fatty acids synthesized in the host plants are transferred to the fungus to sustain

mycorrhizal colonization (Jiang *et al.,* 2017; Bravo *et al.,* 2017; Luginbuehl *et al.,* 2017; Keymer *et al.,* 2017).

The genetic complexity of AMF is also increased by the presence of endobacteria living inside hyphae and spores of many AMF (Bonfante and Desirò, 2017). Two types of endobacteria have been described in AMF: rod-shaped and Gram-negatives, associated to members of the Gigasporaceae family (Ghignone *et al.,* 2012) and coccoid Mollicutes-related endobacteria, distributed across different lineages of AM fungi (Naumann *et al.,* 2010; Desirò *et al.,* 2014; Torres *et al.,* 2015). To date, only *Candidatus* Glomeribacter gigasporarum, the rod-shaped bacterium hosted by *Gigaspora margarita* has been investigated in more detail. Thanks to the comparison with a cured, endobacteria-free, strain (Lumini *et al.,* 2007) it has been demonstrated that the endobacterium enhances fungal sporulation, bioenergetic capacity by priming mitochondrial metabolic pathways and ability to detoxify reactive oxygen species possibly leading to an improved ecological fitness of the fungus (Salvioli *et al.,* 2016; Vannini *et al.,* 2016).

All major fungal lineages were shown to possess additional genetic components among which are viruses also called mycoviruses (Ghabrial *et al.,* 2015; Son *et al.,* 2015). In most cases the genome of mycoviruses is an RNA molecule, mainly a double-stranded RNA (dsRNA) (Fauquet *et al.,* 2005), containing at least a gene encoding an RNA-dependent RNA polymerase (RdRp), which is necessary for replication (Ghabrial and Suzuki, 2009). Mycoviruses are transmitted horizontally *via* hyphal anastomosis between vegetatively compatible individuals or transmitted vertically through asexual and, less frequently, through sexual spores (Hillman *et al.,* 2004; Ghabrial *et al.,* 2015). External route of infection are apparently lacking; however, recent studies suggest that mycophagous insects function as mycovirus transmission vectors in the case of ssDNA mycoviruses (Liu *et al.,* 2016). The infection by mycoviruses is often asymptomatic, but in some specific conditions alterations in the phenotype of the host fungus can be observed (van Diepeningen *et al.,* 2006; Nerva *et al.,* 2017). In some cases higher order biological interactions are

also influenced; for example, the presence of a virus can lead to attenuation of virulence in a plant pathogenic fungus (Choi and Nuss, 1992) or can provide the capability to enhance thermal tolerance of the host plant to an endophytic fungus (Márquez *et al.,* 2007).

The knowledge of mycoviruses in mycorrhizal fungi is still very limited. In the ectomycorrhizal fungus *Tuber aestivum*, three mycoviruses - a mitovirus (Stielow *et al.,* 2011b), a totivirus (Stielow and Menzel, 2010) and an endornavirus (Stielow *et al.,* 2011a) - have been described but their biological role was not characterized. Concerning AMF, Ikeda *et al.,* (2010) demonstrated, for the first time, the presence of mycovirus-related dsRNAs, in the mycelium of *Glomus sp*. (now named *Rhizophagus*) strain RF1. A 4.557 nucleotides segment, called GRF1V-M, which could not be phylogenetically assigned to known genera of mycovirus, was characterized in detail. The GRF1V-M encodes an RdRp and a structural protein. Remarkably, by subculturing single spores, a fungal virus-free line was obtained: the absence of the virus led to the production of a higher number of spores and to an enhanced stimulation of plant growth compared to the GRF1V-M-positive line (Ikeda *et al.,* 2012). In a second work the same group has characterized by deep sequencing of dsRNA in the AMF *Rhizophagus clarus* strain RF1, a sequence showing similarity to RdRp of mitoviruses (Kitahara *et al.,* 2014). In this case a biological function has not been reported.

Here, by exploiting recently published transcriptomic data (Salvioli *et al.,* 2016) we describe for the first time the components of the virome of the AMF *G. margarita*, which is phylogenetically distantly related to *Rhizophagus* species. We provide *in silico* and experimental evidence of the existence of a population of six viral sequences (4 mitoviruses, one Giardia-like and one Ourmia-like virus). With one exception (the Ourmia-like virus), we found that the abundance of the viral RNA molecules was not regulated by the presence of the endobacterium hosted by *G. margarita*. Interestingly, we for the first time report cDNA fragments corresponding to mitovirus sequences associated to the presence of their corresponding viral RNA: these sequences are integrated in neither the mitochondrial nor the nuclear genome, but likely exist as extrachromosomal fragments.

RESULTS

Identification of viral sequences from *Gigaspora margarita* transcriptome

Contigs assembled from RNAseq data for the AMF *Gigaspora margarita* BEG34 (Salvioli *et al.,* 2016) were searched for sequences annotated as RdRp; these are proteins performing enzymatic activities typically encoded by RNA viral genomes, easily distinguishable from host RdRp based on homology to other viral or host (fungal) RdRps. Eight sequences were retrieved and a BLAST search allowed us to identify four mitoviruses, one ourmia-like virus, one Giardia-like dsRNA virus, and two sequences related to Fusarium mycoviruses (Supporting information Table S1 and Fig. S1).

The four putative mitoviral sequences and the Ourmia-like contig encode for a single ORF for each genome, which, based on the presence of the conserved motifs GDD, can be annotated as putative RdRps (Supporting information Fig. S1). The 5' UTRs present a length of 472 bp for 33086, 304 bp for 34036, 281 bp for 34770, 275 bp for 34875 and 268 bp for the Ourmia-like contig. The ORFs encoding the RdRps end at 2854 bp for 33086, 3262 bp for 34036, 3311 bp for 34770, 3299 bp for 34875 and 3061 for the Ourmia-like virus. In the case of the Giardia-like, whose genomic sequence predicts the presence of two ORFs, the first starts at nt 311 and ends at nt 1567 while the second one starts at nt 1611 and ends at nt 4679. Each ORF encodes a putative RdRp protein of 794 aa for 33086, 986 aa for 34036, 1010 aa for 34770, 1008 aa for 34875 and 931 aa for the Ourmia-like virus. The first ORF of the Giardia-like encodes a putative protein of 418 aa with no similarity with proteins currently present in databases; the second ORF of 1022 aa shows similarity to RdRps of another Giardia-like virus isolated in the AMF *Rhizophagus sp.* (Ikeda *et al.,* 2012) (Suppl. Table 1).

To confirm the authenticity (presence as biological entities, i.e. RNA molecules, and not as mere *in silico* artifacts) of the sequences assembled from the transcriptome, RT-PCR assays were performed using as template RNA from *G. margarita* germinating spores with specific primers (Supporting

information Table S2) designed to amplify a cDNA fragment in the size-range of about 500-700 bp. A fragment of the expected size was obtained for each of the four mitoviruses and the Giardia-like virus (Fig. 1). The sequencing of recombinant plasmids, after cloning the PCR fragments confirmed the identity of the sequence of the PCR product with that predicted *in silico* (data not shown). More primer pairs spanning different regions of the *in silico* predicted genomes were designed for the remaining sequences (those that could not be confirmed in the first round of PCR, such as contig 23972, 39980 and 36178, the Fusarium micovirus and the Ourmia-like related sequences) and tested again in RT-PCR assays. No amplified fragment was obtained for the 23972 and 39980 contigs, the two Fusarium mycovirus-related sequences. Only for the Ourmia-like sequence (contig 36178) we could confirm by RT-PCR and sequencing the presence of an RNA molecule corresponding to a short portion (324 bp) identitical to the virus RdRp predicted *in silico* (Supporting information Fig. S2). Although the 36178 contig obtained *in silico* (Ourmia-like putative virus) encodes a long ORF, the amplified segment correspond only to conserved RdRp motifs, whereas the rest of the sequence does not show similarity to existing RdRps; several attempts to verify the presence of other cDNA fragments corresponding to other regions of the Ourmia-like contig extending outside the conserved GDD domain failed: we therefore hypothesized that the *in silico* assembled sequence may be a chimera generated by an artifact of RNAseq assembly, and that only the conserved region we could amplify is corresponding to a specific Ourmia-like RNA molecule.

To confirm the presence of an active RdRp encoded by the different identified sequences, northern blot assays were carried out on *G. margarita* germinating spores using positive and negative strand riboprobes. Northern blot allowed to detect genomic (+sense) RNA corresponding to the four mitoviruses and the narnavirus-like sequence (Ourmia-like contig 36178) (Fig. 2-A-D), but we could provide evidence for active replication only for the mitoviruses since both full length positive strand and negative strand RNA were detected in RNA extracts; in addition, they seem rather more abundant (Fig. 2 A-D) compared to the other virus tested. For the Ourmia-like sequence

only the positive sense-detecting probe gave a hybridization signal; no signal was obtained for the negative sense-detecting probe probably due to the low abundance of the target sequence (Fig. 2-E), close to detection limit of a northern blot assay.

No hybridization signal was detected for the Giardia-like sequence with both probes, suggesting that the abundance of the RNA was below the threshold for northern detection in our experiments (data not shown). The very low abundance of RNA corresponding to the Ourmia-like virus and the Giardia-like virus was confirmed by qRT-PCR assays (see below).

Phylogenetic placement of viral sequences

We then further characterized our viral sequences with a phylogenetic analysis comparing conserved regions of the RdRp, with those present in the databases. We first aligned the *G. margarita* mitovirus and Ourmia-like virus to a dataset of RdRp that comprises the family *Ourmiaviridae*, *Leviviridae*, *Narnaviridae* and a number of related viruses still not classified taxonomically that resulted from NGS virome characterization (Shi *et al*., 2016) The conserved RdRp region among this wide dataset is limited to 109 amino acids surrounding the GDD catalytic motif. This phylogenetic analysis allows to show that *G. margarita* putative mitoviruses are indeed members of the *Mitovirus* genus, whereas the Ourmia-like fragment present in *G. margarita* belongs to a clade of viruses still not classified taxonomically (Supporting information Fig. S3), which we have previously proposed to form a new genus called Ourmia-like viruses (Turina *et al*., 2017).

A further more defined phylogenetic analysis, that includes only mitoviruses, was carried out on aligned sequences spanning 467 amino acids using a plant endogenized mitovirus as outgroup (Fig. 3). This analysis showed that mitoviruses infecting *G. margarita* are phylogenetically diverse, in particular the 33086 is distantly related to the other three sequences which instead constitute a statistically well supported clade (34036, 34770 and 44875) that is distinct from the other well supported mitovirus clades (Fig. 3). Indeed, contig 33086 shares the

highest RdRp protein identity score with *Botrytis cinerea* mitovirus (38.0%), while the best identity value shared with the other *G. margarita* mitoviruses is 30.1% (33086 vs 34875). The other three mitoviruses, instead, share higher percentage identity among each other, respectively from 32.0% (34036 *vs* 34875) to 53.5% (34770 *vs* 34036) (Table 1 and Supporting information Table S3). Interestingly, in all the 4 mitovirus sequences 100% of tryptophan (W) residues are encoded by the TGG codon (Supporting information Fig. S1): therefore, a functional RpRd can be hypothetically translated both in the cytosol and the mitochondria.

A phylogenetic tree for the Giardia-like viral sequence was also generated including the most closely related members of dsRNA virus families (*Totiviridae* and *Partitiviridae*); as expected, virus RdRPs belonging to the *Totiviridae* and *Partitiviridae* are grouped in statistically well supported clades; the rest of the aligned sequences (including the *G. margarita* Giardia-like virus) forms a clade with a relatively low statistical support value (39%) in bootstrap analysis (Fig. 4).

Are the viral sequences endogenized into the *G. margarita* genome?

A widespread endogenization of mitoviral sequences in plant genomes was recently reported (Bruenn *et al.,* 2015) and this prompted us to look for possible genome endogenization events of the viral sequences we identified in *G. margarita.*

We performed PCR assays on DNA extracted from *G. margarita* spores with the same specific primers that were used to amplify the cDNA corresponding to the RNA (Supporting information Table S1). In case of positive outcome, to confirm that the PCR products were indeed originated from DNA, a control sample using DNAse-treated genomic DNA was also analysed in parallel to exclude any spurious activity of the DNA polymerase that might function as reverse transcriptase. A PCR product of the expected size, which was absent in the DNase-treated sample, was obtained for the four mitoviruses and for the Fusarium mycovirus 3-related 23972 sequence (Fig. 5). For these five virus-like sequences a phenomena of genome endogenization can be therefore hypothesized.

To further explore this issue we checked the sequence of the previously characterized *G. margarita* mitochondrial genome (BEG34 strain, the same used in this study) (Pelin *et al.,* 2012): no viral sequence was found within the complete mitochondrial DNA. We then tested the hypothesis that these DNA sequences corresponding to viral RdRps could be endogenized into the nuclear genome. A draft genome sequence of *G. margarita* (BEG 34 strain) is available (Ghignone, Venice, Salvioli, Bonfante, unpublished). We mapped the genome sequencing reads on viral contigs by using BWA software. The DNA sequence encoding the elongation factor, used as a positive control, was well represented (590 paired end reads). We also had a confirmation of the genome endogenization for the 23972 sequence (the Fusarivirus fragment) for which 164 paired end reads were found. Surprisingly, coverage was zero for the 4 mitoviral sequences rejecting the hypothesis of a genomic integration.

To investigate quantitatively the abundance of these viral DNA portions, we performed quantitative PCR assays by using specific primers covering about 100-120 bp fragments of the four mitoviruses (Supporting information Table S2) on three independent DNA preparations from 300 *G. margarita* spores. As endogenous controls we used PCR primers targeting a portion of the ribosomal DNA 18S and of the elongation factor gene. Only the DNA corresponding to the mitovirus 33086 accumulated to a level statistically significantly different from the other three mitoviruses. The amplicon corresponding to 33086 was the least abundant, while that of 34036 was the most abundant but in all cases the amount of viral DNA was statistically significantly lower than that of the DNA sequence encoding the elongation factor, which is supposed to be a single copy gene (Table 2). Even with the more sensitive q-PCR assay we could not detect any DNA fragment related to the Ourmia-like sequence (data not shown). We cloned and sequenced the DNA fragments corresponding to the 34470 and 34875 mitoviruses, and confirmed that the sequence fragments had the same sequence as the one assembled *in silico*, without gaps or any signs of internal recombination (not shown).

To further clarify the nature of the DNA corresponding to the mitovirus sequences, we then investigated whether i) DNA corresponding to the whole virus sequence could be amplified and ii) the size of DNA amplified corresponded to genome fragments or to the complete viral sequence. For this purpose a series of specific PCR primers was designed to span different portions of the genome of the contig 34036 and to obtain amplicons of different size for the same mitoviral sequence 34036 (Fig. 6): we choose the mitovirus 34036 because is the one with the highest DNA titer. The primers were tested on two independent DNA extractions and on two independent cDNA preparations from germinating spores.

We could amplify relatively abundant DNA corresponding to fragments spanning the 3' terminal 2400 bp, and only a much fainter band corresponding to the a fragment at the 5' of the genomic sequence (Fig 6a). The consensus sequence obtained from a number of different clones amplified from DNA corresponded exactly to that predicted from the *in silico* analysis, with only 1 synonimous nucleotide mutation (C to T) at position 1399 (Supporting information Fig. S4). When we tested the maximum length of the amplified segment using different primer combination, independently from the genome fragments where primers were designed, only amplicons below 700 bp in length could be amplified from DNA, while larger PCR products were obtained from cDNA (Fig. 6b). This provides indirect evidence of the existence of small fragments of DNA corresponding to the various regions of most of the viral genome.

The mitoviral sequences are present in *G. margarita* B- (endobacteria free) but not in *Gigaspora rosea* genomic DNA

In order to investigate the presence of homologous viral sequences in other AMF, we performed PCR experiments on genomic DNA from *Gigaspora rosea*, a phylogenetically related AMF which does nor host endobacteria, using the same *G. margarita* virome specific primers (Supporting information Table S2). We also considered DNA of a *G. margarita* BEG34 isolate that has been cured from the *Candidatus* Glomeribacter gigasporarum endosymbiotic bacteria (Lumini *et al.,*

2007) and has been called B- (endobacteria free). The quality of the DNA preparations was first verified by PCR amplifications with AM-specific ribosomal primers AML1 and AML2 (data not shown). A PCR product of the expected size was obtained from the B- isolate for all the analysed viral sequences (4 mitoviruses and the Fusarium mycovirus 3-related sequence). The sequencing of the PCR products confirmed the correspondence to the expected virus sequence (not shown). By contrast, the *G. rosea* sample always gave negative results (Fig. 7). This suggests that these portions of viral DNA are constantly present in *G. margarita* independently from the presence of endobacteria, but are absent from the phylogenetically close *G. rosea*.

Abundance of viral RNA molecules in B+ and B- germinating spores

We then investigated whether the abundance of the viral RNA molecules was regulated by the presence or absence of endobacteria. Quantitative RT-PCR reactions were set up on cDNA obtained from B+ and B- germinating spores using the elongation factor as housekeeping gene. On average, the mitoviral sequences led to a Ct (threshold cycle) between 11 and 14 (Ct =12 for 33086; Ct =14 for 34036; Ct = 11; 34470; Ct = 12 for 34875). For the Ourmia-like virus and the Giardia-like virus sequences an average of 22 and 30 Ct was registered, respectively, confirming the lower abundance compared to the mitoviral RNAs that was already noticed in the northern blot experiments. Remarkably, no statistically significant different virus accumulation between B+ and B- spores was observed for the mitoviral sequences and the Giardia-like virus (Fig. 8) but, surprisingly, the Ourmia-like sequence seems to be present only in B+ spores: Ct values for this virus registered for B- spores were below the detection threshold.

DISCUSSION

A further layer of complexity in the organisms associated to *G. margarita*

The obligate biotrophism and the lack of stable genetic transformation protocols have made AMF a complex biological system to study. Recent works, such as the description of the genome

sequence of *R. irregularis* (Tisserant *et al.,* 2013; Lin *et al.,* 2014), have increased our knowledge on AMF genetic and genomic complexity, and shed light on the molecular determinants that play a key role in the establishment of the symbiosis with the host plant. A further contribute to this complexity is given by the presence of endobacteria (Bonfante and Desirò, 2017), whose biological functions, with a single noticeable exception (Salvioli *et al.,* 2016), are still largely unknown.

A few studies recently revealed that a further level of complexity is given by the presence of mycoviruses, that are widespread in all major taxa of fungi. Mycoviruses in AMF have possibly evolved under unique selection pressures. In one case only it has been demonstrated that they are a biologically active component of the symbiosis: a *Glomus* (*Rhizophagus*) strain free of the GRF1V-M virus (a virus related to the Giardia-like virus described in this work) produced a higher amount of spores and promoted plant growth more efficiently than the fungal line containing the virus (Ikeda *et al.,* 2012). Similarly a fungal strain of *Mortierella elongata* cured from its endobacteria produces higher spore numbers and grows better (Li *et al*., 2017). Nevertheless, virus and endobacterial association with AMF seem to be fairly common in nature, raising the possibility of some ecological advantage difficult to measure in the laboratory experimental conditions, where most of the competition aspects are not under scrutiny. Despite the above-mentioned major acquisition, the knowledge of mycoviruses biological role in AMF is still limited.

Recently published RNAseq data of the AMF *G. margarita* (Salvioli *et al.,* 2016) allowed us to identify through computational analysis, four putative mitoviral sequences, a putative Ourmia-like mycovirus, a putative Giardia-like dsRNA virus and two sequences related to Fusarium mycoviruses. The existence as RNA molecules of *in silico* predicted sequences was confirmed by RT-PCR experiments, for the 4 mitoviral sequences, and the Ourmia-like and Giardia-like sequences. We could not confirm the sequence related to Fusarium mycovirus 1 while that related to Fusarium mycovirus 3 turned out to be corresponding to a stably endogenized viral fragment into the genome of *G. margarita* likely not sufficiently transcribed to be detected by RT-PCR.

We could confirm the presence of the Ourmia-like virus RNA by RT-PCR assays only based on a short sequence of RNA molecule corresponding to a small portion of the sequence (Supporting information Fig. S2) which encodes for an RdRp conserved domain. We hypothesize that the whole 3.2 kb contig is probably a result of a misassembly. It is worth to note that the contig is rich in AT stretches which often give many difficulties during the *in silico* assembly process. Nevertheless, the small segment corresponding to the part amplified by RT-PCR is indeed part of a self replicating RNA molecule, since no DNA corresponding to this fragment was detected, excluding endogenization events. Moreover, we demonstrated that the four mitoviruses are replicative and their RdRP is active since both positive and negative sense RNA strands were detected in northern blot assays. Regarding the Ourmia-like fragment, we obtained a signal only from the positive probe while no signal was detected neither by the positive nor by the negative probe for the Giardia-like, probably due to low RNA abundance. The size of the Ourmia-like RNA detected with the antisense probe (detecting + sense) is approximately 3 kb: therefore indeed an RNA molecule larger than the small fragment we could amplify by RT-PCR exists.

Since many AMF are characterized by the presence of endobacteria which live inside hyphae and spores, and it has been demonstrated that their presence may have an important role for the fungal fitness (Lumini *et al.,* 2007; Salvioli *et al.,* 2016; Vannini *et al.,* 2016), we investigated whether the presence or absence of endobacteria could affect the occurrence and abundance of viral RNA in *G. margarita*. At least in germinating spores, the presence or absence of the endobacteria does not affect the occurrence and the abundance of viral RNA molecules, with the exception of the Ourmia-like virus that seems to be present only in B+ spores. This result opens the possibility that the replication of this virus is dependent on endobacteria likely inside the endobacteria itself. It can be also hypothesized that this sequence is associated to endobacteria; indeed a number of viruses phylogenetically related to Ourmia-like viruses are known phages (family *Leviviridae*) (Dolja and Koonin, 2012).

Phylogenetic analysis of the *G. margarita* virome showed both evidence of virus-host co-evolution and of horizontal gene transfer

Since viral RNA-dependent RNA polymerases (RdRp) are markers of mycoviruses with an RNA genome, we carried out a phylogenetic analysis considering the conserved catalytic domain of the identified RdRps. This first analysis allowed us to associate the four mitoviruses to the genus *Mitovirus* and the Ourmia-like fragment to a new clade called Ourmia-like group that was recently shown to exist after a wide NGS (Next Generation Sequencing) analysis of invertebrate virome (Shi *et al.,* 2016).

A further analysis allowed us to focus on the diversity of the four mitoviruses; in particular, three of them cluster together (showing evidence of co-evolution with their host) while the 33086 sequence is separated and more closely related to mitoviruses from a phylogenetically distant host (*Botrytis cinerea*), providing some indirect evidence of possible horizontal virus transmission. Taxonomically, the criteria to establish a new mitovirus species requires the identity in the whole aligned RdRp to be below 40% (Hillman and Esteban, 2011); when we first checked the best sequences match retrieved by BLAST, none of them shared more than 40% identity, implying that the four mitoviruses constitute new mitovirus species. When we performed a pairwise alignment among the *G. margarita* mitovirus pairs using the Needleman-Wunsch Global Align Protein Sequences software, inside the BLAST suite, none of the mitovirus pairs shared an identity percentage over 40% (39% being the highest, between contig 34470 and 34875) (Table 1). Therefore we propose that each of the mitovirus sequence we detected belongs to a new mitovirus species, respectively called *Gigaspora margarita* mitovirus 1 (GmMV1), GmMV2, GmMV3 and GmMV4.

The Giardia-like sequence also constitutes a new viral species called *G. margarita* Giardia-like virus 1 (GmGLV1). Given that in the case of the Ourmia-like sequence only a small fragment was confirmed by RT-PCR, we think that the data is overall too preliminary to establish a new virus

species. The phylogenetic inference of the Giardia-like sequence shows that RdRp belonging to *Totiviridae* and *Partitiviridae* are grouped in a clade statistically supported while the other aligned sequences, including Giardia-like, show low RdRp sequences identities (39%) and still do not form a statistically well supported clade.

The *G. margarita* mitoviruses display the same TGG codon frequency for tryptophan as the mitochondrial genes

The Mitovirus genus includes fungal viruses with small RNA genomes with a single ORF encoding a RdRp; most mitoviruses replicate in their host's mitochondria (Cole *et al.,* 2000; Hillman and Cai, 2013; Wu *et al.,* 2016). For translation, mitoviruses relies on the endogenous mitochondrial translational code which can use for tryptophan (Trp) the codons TGG (as in the cytosolic/nuclear genetic code) and TGA (which, by contrast, serves as a stop codon in the standard cytosolic/nuclear genetic code). Interestingly, in the four *G. margarita* mitoviral sequences all the Trp residues are encoded by the TGG codon. This observation indicates that, apparently, functional RdRps from the four *G. margarita* mitoviruses can be translated both in cytosol and mitochondria. The use of TGG, compatible with both cytoplasmic and mitocondrial translation, is a feature of a few mitoviruses, including RcMV1-RF1, the only mitovirus described so far in an AMF, that is *Rhizophagus clarus* (Kitahara *et al.,* 2014). It has been hypothesized that this feature might be an advantage for horizontal transmission among AMF (Kithara *et al.,* 2014). However, it has been recently observed that host fungi whose mitoviruses have no or few TGA codons are distinctive in also having no or few TGA codons in their core mitochondrial genes. For example, *G. margarita* mitochondrial DNA has only 2% Trp encoded by TGA and similar low percentages occur in the mitochondrial genomes of other AMF (Nibert, 2017). Thus, the exclusion of such codons in some mitoviruses appears to reflect most fundamentally that TGA is a rare mitochondrial codon in their particular hosts. Whether the four *G. margarita* mitoviruses are capable to replicate also in the cytosol is still an open question and needs to be experimentally proved.

First report of cDNA corresponding to replicating non retroviral RNA viruses present in hosts other than diptera

Since integration events of fungal mitovirus cDNA in the mitochondrial (and to a minor extent nuclear) DNA of vascular plants was shown to be a fairly common event (Bruenn *et al.,* 2015), we investigated whether genome integration occurred also for the viral sequences identified in *G. margarita*. Our PCR assays revealed the presence of DNA templates corresponding to the four mitoviruses and to Fusarium mycovirus 3-related sequence (23972), but when we checked the published sequence of the mitochondrial genome of *G. margarita* we didn't find any viral sequence. The presence of multiple viral sequences in the same individual organism is not uncommon for mycoviruses (Nerva *et al*., 2016; Kondo *et al.,* 2013; Xie and Gahbrial, 2012; Jiang *et al*., 2013), even in the case of closely related viruses as for the 4 *G. margarita* mitovirus species. It would be interesting to assess the exact cellular niche for each mitovirus and whether they all infect each mitochondrium (mixed infection) or whether one mitovirus species for each mitochondria exists, contribuiting to a differentiation among the mitochondrial identities inside a single organism.

We had also the possibility to check for integration of those mitoviral DNA templates in the genomic DNA, looking into reads from an ongoing sequencing project of the same *G. margarita* isolate: we could map reads from the endogenized Fusarium-like virus fragment, but no evidence of endogenized mitovirus sequence could be obtained. The nature of the amplified DNA for mitoviral sequences remains therefore undefined. Our attempts to perform Southern blot assays failed in detecting specific hybridization signals related to mitovirus endogenization (data not shown). Taken as a whole, these data suggest that the mitoviral DNA sequences are not integrated in the genome but they likely exist as extrachromosomal fragments of relatively small size as it can be deduced by our different amplicon length amplification PCR experiments. This is the first time that viral DNA has been detected in the presence of the corresponding homologous replicant RNA for mitoviruses. The presence of genome endogenization in plants has been so far associated with the absence of

replicative capability of mitoviruses (Koonin *et al.,* 2014). To our knowledge this is also the first time that cDNA corresponding to a non retroviral replicating RNA virus is detected outside the phylum arthropoda. In fact a few reports have pointed to the existence of episomic cDNA of RNA viruses in the order Diptera (Nag *et al.,* 2016; Goic *et al.,* 2013; Goic *et al.,* 2016). Recently, it was shown that two species of *Aedes* mosquitoes infected with two arboviruses from distinct families (dengue or chikungunya) generate, by endogenous reverse transcriptase activity, a viral-derived DNA that is essential for mosquito survival and viral tolerance, being at the base of persistent viral infections (Goic *et al.,* 2016); inibition of cDNA synthesis results in higher mortality without affecting the RNAi anti-viral system. In our system, we have shown that cDNA sequences correspond exactly to viral genomic sequences without signs of re-arrangement as it is often the case in the *Drosophila melanogaster* viral infection system (Goic *et al*., 2013). A possible analogy between the two system is that in mosquito it was shown that cDNA was associated to persistent infection and our virus-fungal system also has the hallmarks of persitent infections (lack of external infectivity of mitoviruses, widespread association between these viruses and AMF hosts, lack of obvious detrimental effects on the host, presence of a relatively high virus titer); it is tempting to speculate that the same anti-viral mechanism bringing to persistent infections is common to such distantly related hosts. Furthermore, it would be interesting to identify the mechanisms by which the mitoviral DNAs are generated in *G. margarita* and whether they may have an anti-viral functional role as is the case of mosquito infecting virus *via* the piRNA pathway (reviewed in Olson and Bonizzoni, 2017). Given the large amount of retrotransposon sequences present in the *G. margarita* genome (Ghignone, Venice, Anselem, Salvioli, Bonfante, unpublished), we can envision retro-transcriptase activity that can occasionally use as template RNA mitovirus sequences; the presence of such retroviral activity in mitochondria is, to our knowledge, unknown and should be further tested.

In conclusions, the augmented genome concept that scientists often use to describe human beings at the organismal level, including the whole microbiota genomes, can be in part applied to the fascinating biological system of AMF, where a fungus, its associated bacteria, and the associated viruses seem to constitute an interacting superorganism; understanding the biological relevance of each of these components and how they interact will be the challenging task to be pursued in the near future.

EXPERIMENTAL PROCEDURES

Biological material

Spores of *Gigaspora margarita* Becker and Hall (BEG 34), the corresponding cured strain (without the endobacterium *Candidatus* Glomeribacter gigasporarum; Lumini *et al*., 2007) and *Gigaspora rosea* (BEG 9) were propagated by using white clover (*Trifolium repens*) as trap plant. Clover plants were inoculated with ca. 100 spores and after 2-3 months new spores were generated and collected by the wet sieving technique. To generate germinating spores, *G. margarita* spore suspensions were divided into aliquots of 100, surface sterilized twice for 10 minutes with 3% chloramine-T and 0.03% streptomycin sulfate, rinsed several times with sterile distilled water and then incubated in 1 ml of sterile distilled water for 5-7 days in the dark at 30°C. Germinated spores were collected, immediately frozen in liquid nitrogen and stored at -80°C.

Total nucleic acids extraction

About 200-300 spores were crushed by a pestel using 1 ml of lysis CTAB buffer containing 2% cetyl trimethylammonium bromide, 1% polyvinyl pyrrolidone, 100 mM Tris-HCl, 1.4 M NaCl, 20 mM EDTA. After incubation at 65°C for 15 min, samples were centrifuged at 10,000 rpm for 10 min. The supernatant was added with an equal volume of phenol-chlorophorm-isoamyl alcohol (25:24:1; vol:vol:vol) and mixed. The upper acqueous phase of a centrifugation at 10,000 rpm for 10 min was transferred into a new tube to which an equal volume of chlorophorm was added. The

aqueous phase was then collected and precipitated with 2/3 vol of isopropanol. After an incubation at 4°C for 2 hrs, semples were centrifugated at 14,000 for 30 min at 4°C. The pellet was washed with 80% ethanol and resuspended into 50 µl of sterile water. To confirm that the PCR products were indeed originated from DNA, a control sample was generated by a DNAse treatment performed on 1 µg of total nucleic acids for 1 h at 37°C using DNase (Ambion) according to manufactures's instructions.

PCR and cloning

PCR reactions were set up following standards procedures. For sequencing the high fidelity Phusion DNA polymerase (ThermoFisher) was used. Primers are shown in Supporting information Table S2. PCR products were cloned into the TOPO Vector (Invitrogen) vector following manufactures's instructions. Recombinant DNAs were extracted with QIAGEN plasmid Minikit and sequences were obtained form the Sequencing Service, LMU Biozentrum, Großhaderner, Germany). Sequence analyses were performed with CHROMAS LITE (http://www.technelysium.com.au/chromas_lite.html).

RNA extractions and RT-PCR

Total RNA was extracted using the RNeasy Microarray Tissue Mini Kit (Qiagen, Hilden, Germany), according to the manufacturer's instructions. The concentration and quality of the nucleic acids were assessed with a Nanodrop1000 (Thermo Scientific, Wilmington, NC, USA).

Samples were treated with TURBO DNase (Ambion) according to the manufacturer's instructions. The RNA samples were routinely checked for DNA contamination by means of RT-PCR analysis, using primers Efgig2F 5'-TGAACCTCCAACCAGACCAACTG-3' and EfgigR 5'-CGGTTTCAACACGACCTACAGGGAC-3' for *G. margarita* translation elongation factor (Efgig, Salvioli *et al.,* 2014) and the One-Step RT-PCR kit (Qiagen). For single-strand cDNA synthesis samples were denatured at 65°C for 5 min and then reverse-transcribed at 25°C for 10 min, 42°C for 50 min and 70° for 15 min in a final volume of 20 µl containing 10 µM random

hexamers, 0.5 mM dNTPs, 4 µl 5X buffer, 2 µl 0.1 M DTT, and 1 µl Super-ScriptII (Invitrogen). Quantitative RT-PCR (qRT-PCR) experiments were carried out in a final volume of 10 µl containing 5 µl of iTaq™ Universal SYBR® Green Supermix (Bio-Rad), 0.2 µl of 2 µM specific primers (Table 2), and about 20 ng of cDNA. Samples were run in the iCycler iQ apparatus (Bio-Rad) using the following program: 3 min pre-incubation at 95 °C, followed by 40 cycles of 10 s at 95 °C, and 30 sec at 60 °C. Each amplification was followed by melting curve analysis (60 °C to 94 °C) with a heating rate of 0.5 °C every 15 s. All reactions were performed on at least three biological and three technical replicates and only Ct values with a standard deviation that did not exceed 0.3 were considered. The comparative threshold cycle method (Rasmussen, 2001) was used to calculate relative expression levels using as a reference gene for transcript normalization the *G. margarita* elongation factor (Efgig). Statistical analyses were performed through one-way analysis of variance (one-way ANOVA) and Tukey's post hoc test, using a probability level of $p<0.05$. All statistical analyses were performed using the PAST statistical package (version 2.16; Hammer *et al.,* 2001).

Northern blot

For northern blot analyses, total RNA from fungi was prepared using Total Spectrum RNA Reagent (Sigma-Aldrich, Saint Louis, MO, USA) as suggested by the manufacturer. RNA samples were separated in gel electrophoresis under denaturing conditions using glyoxal in HEPES-EDTA buffer as detailed in Sambrook et al. (1989). Hybridization were performed using a radio-labeled RNA probe prepared from EcoRI linearized purified plasmids containing cDNA fragments corresponding to both orientation (sense and antisense probes) of each of the 6 viruses through T7 transcription using the Maxiscript T7 kit reagents (Thermo Fisher Scientific Inc., Waltham, MA, USA), as detailed before (Nerva *et al.,* 2017). In some cases, some fragments in specific orientations were toxic to *E. coli* after transformation, and recombinant plasmid could not be obtained. The problem was circumvented using as template for transcription, instead of a linearized plasmid, a PCR

525 fragment amplified with T7 promoter primer, and a fragment specific reverse primer directly
526 amplified from the ligation reaction.

527 **Phylogenetic analyses**

528 We identified groups of conserved sequences related to the different mycoviruses present in *G.*
529 *margarita* through BLAST searches of the databases (Accession numers reported in Supporting
530 information Table S4) and the representative protein sequences identified were used for multiple
531 sequence alignments using MUSCLE (Edgar, 2004) implemented in the MEGA 6 (Tamura *et al*.
532 2013). Aligned sequences were used for infer phylogenetic trees using the Maximum Likelihood
533 method based on the Le_Gascuel_2008 model (Le and Gascuel 2008). The best amino acid
534 substitution model was calculated with MEGA 6. Statistical analysis was carried out through
535 bootstrap analysis with 1000 replicates. Further details of the phylogenetic analysis are included in
536 figure legends. Multiple aligned sequences were also used to calculate pairwise identity and
537 similarity percentages using MatGat (Campanella *et al*., 2003).

538 **Bioinformatic analyses**

539 *G. margarita* BEG34 genomic DNA libraries (PE, MP-3kb, MP-8kb), currently used for an ongoing
540 genome assembly project, were first checked for quality with FASTQC (Andrews, 2010) and then
541 trimmed with TRIM GALORE! (Krueger, 2012). Cleaned PE reads were mapped onto *G.*
542 *margarita* transcript comp11141_c0_seq1 (GBYF01010162.1), coding for Translation Elongation
543 Factor EF-1 alpha, and onto putative viral sequences (33086, 34036, 34470, 34875) using BWA (Li
544 and Durbin, 2009). Mapping outputs were handled and analyzed with SAMTOOLS (Li *et al.,*
545 2009). Clues of endogenization of the putative Fusarium graminearum dsRNA mycovirus-3-like
546 sequence in the *G. margarita* genome where searched querying the assembly with the
547 comp23972_c0_seq1 (GBYF01024012) sequence using BLASTN 2.6.0+ (Zhang *et al.,* 2000), and
548 mapped reads were counted after BWA analysis.

549 The viral sequences have been submitted to GenBank under the following accession numbers:
550 33086: MG256173, 34036: MG256174, 34470: MG256175, 34875: MG256176 and 33452:
551 MG256177.

552

553 **ACKNOWLEDGEMENTS**

554 The research was supported by the 60% Projects (University of Torino) to PB and LL. The Authors
555 thank Dr. Mara Novero for the technical assistance and Dr. Simone Belmondo for its contribution in
556 the first phases of the project.

557

558

559

REFERENCES

Andrews, S. (2010) FastQC: a quality control tool for high throughput sequence data. URL: http://www.bioinformatics.babraham.ac.uk/projects/fastqc

Berruti, A., Lumini, E., Balestrini, R., and Bianciotto, V. (2015) Arbuscular mycorrhizal fungi as natural biofertilizers: let's benefit from past successes. Front Microbiol 6: 1559.

Bonfante, P., and Desirò, A. (2017) Who lives in a fungus? The diversity, origins and functions of fungal endobacteria living in Mucoromycota. ISME J 11: 1727-1735.

Bravo, M., Brands, V., Wewer, P., Dörmann, P. and Harrison, M.J. (2017) Arbuscular mycorrhiza-specific enzymes FatM and RAM2 fine-tune lipid biosynthesis to promote development of arbuscular mycorrhiza. New Phytol 214: 1631-1645.

Bruenn, J.A., Warner, B.E., Yerramsetty, P. (2015) Widespread mitovirus sequences in plant genomes. PeerJ. 3:e876.

Campanella, J.J., Bitincka, L. and Smalley, J. (2003) MatGAT: An application that generates similarity/identity matrices using protein or DNA sequences. BMC Bioinform 4.

Choi, G.H. and Nuss, D.L. (1992) A viral gene confers hypovirulence-associated traits to the chestnut blight fungus. EMBO J 11: 473-477.

Cole, T.E., Hong, Y., Brasier, C.M. and Buck, K.W. (2000) Detection of an RNA-dependent RNA polymerase in mitochondria from a mitovirus-infected isolate of the Dutch Elm disease fungus, *Ophiostoma novo-ulmi.* Virology 268: 239-243.

Corradi, N., and Brachmann, A. (2016) Fungal mating in the most widespread plant symbiosis? Trends in Plant Sci 22: 175-183.

Desirò, A., Salvioli, A., Ngonkeu, E.L., Mondo, S.J., Epis, S., Faccio, A. *et al.* (2014) Detection of a novel intracellular microbiome hosted in arbuscular mycorrhizal fungi. ISME J 8: 257-270.

Dolja, V.V., and Koonin, E.V. (2012) Capsid□Less RNA Viruses. In: eLS. John Wiley & Sons Ltd, Chichester. URL: http://www.els.net

Edgar, R.C. (2004) MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res 32(5): 1792-1797.

Fauquet, C.M., Mayo, M.A., Maniloff, J., Desselberger, U., and Ball, L.A. (2005) Virus taxonomy: the eighth report of the International Committee on Taxonomy of Viruses. Amsterdam, NL: Elsevier Academic Press.

Ghabrial, S.A., Castón, J.R., Jiang, D., Nibert, M.L., Suzuki N. (2015) 50-plus years of fungal viruses. Virology 479-480: 356-68.

Ghabrial, S.A., and Suzuki, N. (2009) Viruses of plant pathogenic fungi. Annu Rev Phytopathol 47: 353-384

610 Ghignone, S., Salvioli, A., Anca, I., Lumini, E., Ortu, G., Petiti, L. *et al.* (2012) The genome of the
611 obligate endobacterium of an AM fungus reveals an interphylum network of nutritional interactions.
612 ISME J 6: 136-145.
613
614 Goic, B., Vodovar, N., Mondotte, J.A., Monot, C., Frangeul, L., Blanc, H. *et al.* (2013) RNA-
615 mediated interference and reverse transcription control the persistence of RNA viruses in the insect
616 model *Drosophila*. Nat Immunol 14: 396-403.
617
618 Goic, B., Stapleford, K. A., Frangeul, L., Doucet, A. J., Gausson, V., Blanc, H., *et al.* (2016)Virus-
619 derived DNA drives mosquito vector tolerance to arboviral infection. Nat Commun *7:* 12410.
620
621 Jiang, Y., Wang, W., Xie, Q., Liu, N., Liu, L., Wang, D. *et al.* (2017) Plants transfer lipids to
622 sustain colonization by mutualistic mycorrhizal and parasitic fungi. Science 356: 1172-1175.
623
624 Jiang, D., Fu, Y., Li, G. and Ghabrial, S.A. (2013) Viruses of the Plant Pathogenic Fungus
625 *Sclerotinia sclerotiorum*. In: Ghabrial SA, editor. Adv Virus Res 86: 215-248.
626
627 Hillman, B.I. and Cai, G. (2013) The family Narnaviridae: simplest of RNA viruses. Adv Virus Res
628 86: 149-176.
629
630 Hillman, B.I, and Esteban, R. (2011) Narnaviridae, In: King AMQ, Adams MJ, Carstens EB,
631 Lefkowitz EJ. (Eds.),Virus taxonomy: classification and nomenclature of viruses: Ninth Report of
632 the International Committee on Taxonomy of Viruses. Elsevier, San Diego, 2011. pp.1055-1060.
633
634 Hillman, B.I., Supyani, S., Kondo, H. and Suzuki, N. (2004) A reovirus of the fungus
635 *Cryphonectria parasitica* that is infectious as particles and related to the coltivirus genus of animal
636 pathogens. J Virol 78: 892-898.
637
638 Ikeda, Y., Shimura, H., Kitahara, R., Masuta, C., and Ezawa, T. (2012). A novel virus-like double-
639 stranded RNA in an obligate biotroph arbuscular mycorrhizal fungus: a hidden player in
640 mycorrhizal symbiosis. Mol Plant-Microbe Interact 25: 1005-1012.
641
642 Kamel, L., Keller-Pearson, M., Roux, C., and Ané, J.-M. (2016) Biology and evolution of
643 arbuscular mycorrhizal symbiosis in the light of genomics. New Phytol 213: 531-536.
644
645 Keymer, A., Pimprikar, P., Wewer, V., Huber, C., Brands, M., Simone L Bucerius, S *et al.* (2017)
646 Lipid transfer from plants to arbuscular mycorrhiza fungi. eLife 6, e29107.
647
648 Kitahara, R., Ikeda, Y., Shimura, H., Masuta, C. and Ezawa T. (2014). A unique mitovirus from
649 Glomeromycota, the phylum of arbuscular mycorrhizal fungi. Arch Virol 159:2157-2160.
650
651 Kondo, H., Kanematsu, S. and Suzuki, N. (2013) Viruses of the White Root Rot Fungus, Rosellinia
652 necatrix. In: Ghabrial SA, editor. Adv Virus Res 86: 177-214.
653
654 Koonin, E.V., and Dolja, V.V. (2014) Virus world as an evolutionary network of viruses and
655 capsidless selfish elements. Microbiol Mol Biol Rev 78: 278-303.
656
657 Krueger, F. (2012) A wrapper tool around Cutadapt and FastQC to consistently apply quality and
658 adapter trimming to FastQ files, with some extra functionality for MspI-digested RRBS-type

(Reduced Representation Bisufite-Seq) libraries. URL: https://www.bioinformatics.babraham.ac.uk/projects/trim_galore/

Le, S.Q. and Gascuel, O. (2008) An improved general amino acid replacement matrix. Mol Biol Evol 25:1307-1320.

Liu, S., Xie, J., Cheng, J., Li, B., Chen, T., Fu, Y. *et al.* (2016) Fungal DNA virus infects a mycophagous insect and utilizes it as a transmission vector. Proc. Natl. Acad. Sci. USA 113: 12803-12808.

Li, H., and Durbin, R. (2009) Fast and accurate short read alignment with Burrows-Wheeler Transform. Bioinformatics 25: 1754-60.

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N. *et al.* (2009) The Sequence alignment/map (SAM) format and SAMtools. Bioinformatic*s* 25: 2078-2079.

Li, Z., Yao, Q.M., Dearth, S.P., Entler, M.R., Castro Gonzalez, H.F., Uehling, J.K. *et al.* (2017) Integrated proteomics and metabolomics suggests symbiotic metabolism and multimodal regulation in a fungal-endobacterial system. Env Microb 19:1041-1053.

Lin, K., Limpens, E., Zhang, Z.H., Ivanov, S., Saunders, D.G.O., Mu, D.S. *et al.* (2014) Single nucleus genome sequencing reveals high similarity among nuclei of an endomycorrhizal fungus. PLoS Genet 10:e1004078.

Luginbuehl, L.H., Menard, G.N., Kurup, S., Van Erp, H., Radhakrishnan, G.V., Breakspear, A., Oldroyd, G.E.D. and Eastmond, P.J. (2017) Fatty acids in arbuscular mycorrhizal fungi are synthesized by the host plant. Science 356: 1175-1178.

Lumini, E., Bianciotto, V., Jargeat, P., Novero, M., Salvioli, A., Faccio, A., Bécard, G. and Bonfante, P. (2007) Presymbiotic growth and sporal morphology are affected in the arbuscular mycorrhizal fungus *Gigaspora margarita* cured of its endobacteria. Cellular Microbiology 9: 1716-1729.

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

Nag, D.K., Brecher, M. and Kramer, L.D. (2016) DNA forms of arboviral RNA genomes are generated following infection in mosquito cell cultures. Virology 498: 164-171.

Naito, M,, Morton, J.B., and Pawlowska, T.E. (2015) Minimal genomes of mycoplasma-related endobacteria are plastic and contain host-derived genes for sustained life within Glomeromycota. Proc Natl Acad Sci USA 112: 7791-7796.

Naumann, M., Schuβler, A., Bonfante, P. (2010) The obligate endobacteria of arbuscular mycorrhizal fungi are ancient heritable components related to the Mollicutes. ISME J 4: 862-871.

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 Res 219: 22-38.

709 Nerva L, Silvestri A, Ciuffo M, Palmano S, Varese GC, Turina M. (2017) Transmission of
710 *Penicillium aurantiogriseum* partiti-like virus 1 to a new fungal host (*Cryphonectria parasitica*)
711 confers higher resistance to salinity and reveals adaptive genomic changes. Env Microb 19: 4480-
712 4492.
713
714 Nibert, M.L. (2017) Mitovirus UGA(Trp) codon usage parallels that of host mitochondria.
715 Virology 507: 96-100.
716
717 Olson, K.E. and Bonizzoni, M. (2017) Nonretroviral integrated RNA viruses in arthropod vectors:
718 an occasional event or something more? Curr Opin Insect Sci 22: 45-53.
719
720 Pelin, A., Pombert, J.F., Salvioli, A., Bonen, L., Bonfante, P., and Corradi, N. (2012) The
721 mitochondrial genome of the arbuscular mycorrhizal fungus *Gigaspora margarita* reveals two
722 unsuspected transsplicing events of group I introns. New Phytol 194: 836-845.
723
724 Petrzik K., Sarkisova, T., Starý, J., Koloniuk, I., Hrabáková, L. and Kubešová, O. (2016) Molecular
725 characterization of a new monopartite dsRNA mycovirus from mycorrhizal *Thelephora terrestris*
726 *(*Ehrh.) and its detection in soil oribatid mites (Acari: Oribatida). Virology 489: 12-19.
727
728 Rasmussen, R. (2001) Quantification on the LightCycler. In S Mener, C Wittwer, K Nakagawara,
729 eds, Rapid Cycle Real-Time PCR: Methods and Applications. Springer Press, Heidelberg, pp 21-34.
730
731 Rodriguez, A. and Sanders, I. R. (2015) The role of community and population ecology in applying
732 mycorrhizal fungi for improved food security. ISME J 9: 1053-1061.
733
734 Ropars, J., Toro, K.S., Noel, J, Pelin, A., Charron, P., Farinelli, L., *et al.* (2016) Evidence for the
735 sexual origin of heterokaryosis in arbuscular mycorrhizal fungi. Nat Microbiol 21: 16033.
736
737 Salvioli, A., Ghignone, S., Novero, M., Navazio, L., Venice, F., Bagnaresi, P. and Bonfante, P.
738 (2016) Symbiosis with an endobacterium increases the fitness of a mycorrhizal fungus, raising its
739 bioenergetic potential. ISME J 1-15.
740
741 Sambrook, J., Fritsch, E.F. and Maniatis T. (1987) Molecular cloning: a laboratory manual. - 2nd
742 ed. Cold Spring Harbor, N.Y. Cold Spring Harbor Laboratory Press.
743
744 Sędzielewska Toro, K. and Brachmann, A. (2016) The effector candidate repertoire of the
745 arbuscular mycorrhizal fungus *Rhizophagus clarus*. BMC Genomics 17:101.
746
747 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.,
748 Buchmann, J., Wang, W. Xu, J., Holmes, E.C., and Zhang, Y.Z. (2016) Redefining the invertebrate
749 RNA virosphere. *Nature* 540: 539-543.
750
751 Son, M., Ju, J., and Kim, K-H. (2015) Five questions about mycoviruses. PLoS Pathog 11:
752 e1005172.
753
754 Spatafora, J.W., Chang, Y., Benny, G.L., Lazarus, K., Smith, M.E., Berbee, M.L. *et al.* (2016) A
755 phylum-level phylogenetic classification of zygomycete fungi based on genome-scale data.
756 Mycologia 108: 1028-1046.
757

Stielow, B., Klenk, H.P., Winter, S., and Menzel, W. (2011) A novel *Tuber aestivum* (Vittad.) mitovirus. Arch Virol 156: 1107-1110.

Stielow, B., Klenk, H.P., Winter, S., and Menzel, W. (2011) Complete genome sequence of the first endornavirus from the ascocarp of the ectomycorrhizal fungus *Tuber aestivum* Vittad. Arch Virol 156: 343-345.

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

Tamura, K., Stecher, G., Peterson, D., Filipski, A., Kumar, S. (2013) MEGA6: Molecular Evolutionary Genetics Analysis Version 6.0. Mol Biol Evol 30: 2725-2729.

Tang, N., San Clemente, H., Roy, S., Bécard, G., Zhao, B., and Roux, C. (2016) A survey of the gene repertoire of *Gigaspora rosea* unravels conserved features among Glomeromycota for obligate biotrophy. Front Microbiol 7:233.

Tisserant, E., Malbreil, M., Kuo, A., Kohler, A., Symeonidi, A., Balestrini, R., *et al.* (2013) Genome of an arbuscular mycorrhizal fungus provides insight into the oldest plant symbiosis. Proc Natl Acad Sci USA 110: 20117-20122.

Torres-Cortés, G., Ghignone, S., Bonfante, P., and Schüßler, A. (2015) Mosaic genome of endobacteria in arbuscular mycorrhizal fungi: transkingdom gene transfer in an ancient mycoplasma-fungus association. Proc Natl Acad Sci USA 112: 7785-7790.

Turina, M., Hillman, B.I., Izadpanah, K., Rastgou, M., Rosa, C., Consortium, I.R. (2017) ICTV Virus Taxonomy Profile: Ourmiavirus. J Gen Virol 98: 129-130.

van Diepeningen, A.D., Debets, AJ., and Hoekstra, R.F. (2006) Dynamics of dsRNA mycoviruses in black Aspergillus populations. Fungal Genet Biol 43: 446-452.

Vannini, C., Carpentieri, A., Salvioli, A., Novero, M., Marsoni, M., Testa, L. *et al.* (2016) An interdomain network: the endobacterium of a mycorrhizal fungus promotes antioxidative responses in both fungal and plant hosts. New Phytol 211: 265-275.

Wu, M., Deng, Y., Zhou, Z., He, G., Chen, W. and Li, G. (2016) Characterization of three mycoviruses coinfecting the plant pathogenic fungus *Sclerotinia nivalis*. Virus Res 223: 28-38.

Xie, J., Ghabrial, S.A. (2012) Molecular characterizations of two mitoviruses co-infecting a hyovirulent isolate of the plant pathogenic fungus *Sclerotinia sclerotiorum*. Virology 428: 77-85.

Yaegashi, H., Nakamura, H., Sawahata, T., Sasaki, A., Iwanami, Y., Ito, T., and Kanematsu, S. (2013) Appearance of mycovirus-like double-stranded RNAs in the white root rot fungus, *Rosellinia necatrix*, in an apple orchard. FEMS Microbiol Ecol 83: 49-62.

Young, J.P.W. (2015) Genome diversity in arbuscular mycorrhizal fungi. *Curr Opin Plant Biol* 23: 113-119.

Zhang, Z., Schwartz, S., Wagner, L., and Miller, W. (2000) A greedy algorithm for aligning DNA sequences. J Comput Biol 7: 203-214.

TABLES

Table 1. Percentages similarity (italics) and identity (underlined) of pairwise alignments of the four *G. margarita* mitovirus sequences along the entire protein sequence. Identity and similarity among pairwise alignments were calculated using the Needleman-Wunsch Global Align Protein Sequence algorithm.

Mitovirus contigs	**33086**	**34036**	**34470**	**34875**
33086		*34%*	*32%*	*33%*
34036	21%		*38%*	*40%*
34470	19%	23%		*56%*
34875	20%	24%	39%	

Table 2. Ct values (+/- standard deviation) registered by qPCR on *G. margarita* genomic DNA (3 independent samples) with primers for the 4 mitoviral sequences. 18S rDNA and *Tef* were used as endogenous genes corresponding to multiple or single copy genes, respectively. Different letters indicate statistically significant differences (ANOVA, $p < 0.05$).

	33086	**34086**	**34470**	**34875**	**Tef**	**18S rDNA**
Ct	30.49	27.72	28.46	29.16	25.37	20.93
standard deviation	+/- 0.6647	+/-0.2121	+/-0.0141	+/-0.1273	+/-0.1697	+/-0.3111
Tukey test	d	b	bc	c	a	e

FIGURE LEGENDS

Fig. 1. Gel electrophoresis of RT-PCR products amplified from cDNA from *G. margarita* germinating spores with the following primer pairs specific for the putative viral sequences: 33086genF1 and 33086genR1, 34036Fgen and 34036Rgen, 34770Fgen and 34770Rgen, 34875Fgen and 34875Rgen, 36178Fgen and 36178Rgen, 33452Fgen and 33452Rgen, 23972Fgen and 23972Rgen, 39980Fgen and 39980Rgen. M: 100 bp ladder (Invitrogen); no cDNA sample (-).

Fig. 2. Northern blot analysis of *G. margarita*-associated viral sequences. Time of autoradiography exposure (Exp) is indicated on top of each panel. MUT4330 (total RNA from an unrelated filamentous fungus) was used as negative control in the hybridization experiment. Lower panels are methylene blue stained membranes to show ribosomal RNA loading. gRNA= genomic RNA; rRNA=ribosomal RNA.

Fig. 3. Molecular phylogenetic analysis by Maximum Likelihood method of Mitovirus sequences (*G. margarita* sequences are indicated by arrows). The tree with the highest log likelihood (-22958.4232) is shown. Initial tree(s) for the heuristic search were obtained by applying the Neighbor-Joining method to a matrix of pairwise distances estimated using a JTT model. A discrete Gamma distribution was used to model evolutionary rate differences among sites: 5 categories (+G, parameter = 1.0199). The rate variation model allowed for some sites to be evolutionarily invariable ([+I], 3.8500% sites). The tree is drawn to scale, with branch lengths measured in the number of substitutions per site. All positions with less than 95% site coverage were eliminated. There were a total of 467 positions in the final dataset.

Fig. 4. Molecular Phylogenetic analysis by Maximum Likelihood method of the Giardia-like viral sequence in the context of some representative of the established family taxa *Totiviridae* and *Partitiviridae*. The tree with the highest log likelihood (-11851.6550) is shown. The percentage of trees in which the associated taxa clustered together is shown next to the branches. Initial tree(s) for the heuristic search were obtained automatically by applying Neighbor-Join and BioNJ algorithms

857 to a matrix of pairwise distances estimated using a JTT model, and then selecting the topology with
858 superior log likelihood value. The rate variation model allowed for some sites to be evolutionarily
859 invariable ([+I], 3.5294% sites). The tree is drawn to scale, with branch lengths measured in the
860 number of substitutions per site. The analysis involved 29 amino acid sequences. All positions
861 containing gaps and missing data were eliminated. There were a total of 170 positions in the final
862 dataset. The star indicates the position of the *G. margarita* Giardia-like virus characterized in this
863 study.

864 **Fig. 5**. Gel electrophoresis of PCR products of *G. margarita* DNA untreated (1) or treated (2) with
865 DNAse with primers specific for the viral sequences (the same primer pairs used in figure 1).
866 Samples 3 correspond to reactions with no DNA. M: 100 bp DNA marker (Invitrogen).

867 **Fig. 6**. Characters of the DNA corresponding to mitovirus sequences. A: Gel electrophoresis of
868 PCR products obtained from *G. margarita* cDNA (1), DNA extractions after DNAse (2) and DNA
869 extractions (3); B: Gel electrophoresis of PCR products obtained from two *G. margarita*
870 independent genomic DNA extractions (1, 2), two independent cDNAs preparations (3, 4) or no
871 DNA (5) samples using the different primer pairs. The size of the amplicon in base pairs (bp) is
872 indicated in brackets. C: Scheme of the 34036 viral sequence showing the position of primers
873 (F=forward; R=reverse) indicated by arrows.

874 **Fig. 7**. Gel electrophoresis of PCR products from genomic DNA samples: *G. margarita* B+ (1) and
875 B- (2) harbouring endobacteria or endobacteria-cured, respectively; *G. rosea* (3); *G. margarita*
876 positive control (4); no DNA (5). The specific primer pair used for each virus is listed in the legend
877 of Figure 1.

878 **Fig. 8**. Quantitative RT-PCR on cDNA of B+ and B- (harbouring endobacteria or endobacteria
879 cured, respectively) germinating spores of *G. margarita* with primers specific for the 4 mitovirus
880 and the Giardia-like virus. *GmTef* was used as housekeeping gene and B- sample as reference. No

statistically significant difference was found between B+ and B- samples for any of the five viruses displayed (ANOVA; $p<0.05$).

SUPPORTING INFORMATION

Fig. S1. Nucleotide sequences of the full-length contigs assembled in silico corresponding to the virus sequences in fasta format. RdRp start and stop codons are indicated in bold and cyan. When relevant, the corresponding translation frame is given in order to outline the codons encoding for tryptophan (W highlighted in yellow). Amino acid sequence of the RdRp are also then reported in fasta format and the GDD conserved motif is indicated in pink.

Fig. S2. Gel electrophoresis of PCR products amplified from *G. margarita* genomic DNA (1); *G. margarita* genomic DNA treated with DNase (2); *G. margarita* cDNA from germinating spores (3); no DNA sample (4) with primers specific for the Ourmia-like sequence (36178 contig). M: 1 kb ladder (Invitrogen).

Fig. S3. Molecular Phylogenetic analysis by Maximum Likelihood method of the four mitoviruses and the Ourmia virus (positions in the tree pointed by red arrows). The evolutionary history was inferred by using the Maximum Likelihood method based on the Le_Gascuel_2008 model (Le and Gascuel, 2008). The tree with the highest log likelihood (-11687.6554) is shown. The percentage of trees in which the associated taxa clustered together is shown next to the branches. Initial tree(s) for the heuristic search were obtained automatically by applying Neighbor-Join and BioNJ algorithms to a matrix of pairwise distances estimated using a JTT model, and then selecting the topology with superior log likelihood value. A discrete Gamma distribution was used to model evolutionary rate differences among sites (5 categories (+G, parameter = 1.0325)). The rate variation model allowed for some sites to be evolutionarily invariable ([+I], 2.9115% sites). The tree is drawn to scale, with branch lengths measured in the number of substitutions per site. The analysis involved 64 amino acid sequences. All positions containing gaps and missing data were

906 eliminated. There were a total of 109 positions in the final dataset. Evolutionary analyses were
907 conducted in MEGA6 (Tamura *et al.,* 2013).

908 **Fig. S4.** Schematic representation of the assembly of the clones used for deriving the 2kb consensus
909 sequence corresponding to 34036 mitovirus sequence amplified by PCR using as template DNA
910 preparations (compare with Fig. 6). In blue the amplified genomic region. Vertical green bars
911 correspond to single nucleotide mutations in specific position in at least one of the cloned
912 amplicons in respect to the consensus sequence. The only conserved mutation compared to the *in*
913 *silico* assembled sequence from transcriptome is a synonymous C to T change at position 1399.
914 Gray arrows represent the full-length viral contig sequence (derived from transcriptome, top arrow)
915 and all the different sequences derived from cloning the PCR amplification products displayed in
916 Fig. 6.

917

918 **Table S1**: BLAST searches of databases using as query the viral RdRp contigs present in *G.*
919 *margarita* transcriptome.

920 **Table S2.** List of primers used in this study.

921 **Table S3**: Percentage identity (upper right triangle) and similarity (lower left triangle) among
922 aligned RdRp from different mitoviruses.

923 **Table S4**: Accession numbers and virus names of proteins used in the phylogenetic analysis.