Identification of viruses from fungal transcriptomes

Yeonhwa Jo^{1, #}, Hoseong Choi^{2, #}, Hyosub Chu³, and Won Kyong Cho^{1, *}

¹Research Institute of Agriculture and Life Sciences, College of Agriculture and Life Sciences, Seoul National University, Seoul 08826, Republic of Korea

²Department of Agricultural Biotechnology, College of Agriculture and Life Sciences, Seoul National University, Seoul 08826, Republic of Korea

³Core Protein Resources Center, Daegu Gyeongbuk Institute of Science and Technology, Daegu 42988, Republic of Korea

[#]These authors contributed equally to this work.

*Corresponding author

Dr. Won Kyong Cho

Email: wonkyong@gmail.com

1 **ABSTRACT**

2

3 Viruses infecting fungi are referred to as mycoviruses. Here, we carried out *in silico*
4 mycovirome studies using public fungal transcriptomes. We identified 468 virus-associated
5 contigs assigned to five orders, 21 families, 26 genera, and 88 species. We assembled 120
6 viral genomes with diverse RNA and DNA genomes. The phylogenetic tree and genome
7 organization unveiled the possible host origin of mycovirus species and diversity of their
8 genome structures. Most identified mycoviruses were originated from fungi; however, some
9 mycoviruses had strong phylogenetic relationships with those from insects and plants. The
10 viral abundance and mutation frequency of mycoviruses were very low; however, the
11 compositions and populations of mycoviruses were very complex. Although coinfection of
12 diverse mycoviruses in the fungi was common in our study, most mycoviromes had a
13 dominant virus species. The compositions and populations of mycoviruses were more
14 complex than we expected. *Monilinia* viromes revealed that there were strong deviations in
15 the composition of viruses and viral abundance among samples. *Gigaspora* viromes showed
16 that the chemical strigolactone might promote virus replication and mutations, while
17 symbiosis with endobacteria might suppress virus replication and mutations. This study
18 revealed the diversity and host distribution of mycoviruses.

19

20 **KEYWORDS:** Fungus, mutation, mycovirus, mycovirome, phylogenetics, replication,
21 transcriptome, virome, virus

1 INTRODUCTION

2 Viruses are small infective agents composed of nucleic acids including DNA or RNA that
3 replicate in the infected host (1). Viruses are ubiquitous, presenting not only in most living
4 organisms such as animals, plants, fungi, and bacteria but also in diverse environments (2-6).

5 Viruses do not have a conserved genetic element like bacteria and fungi (7).
6 Moreover, isolation of pure viral particles is very difficult and challenging for the viruses
7 with low titers and coinfecting viruses (8). The rapid development of next-generation
8 sequencing (NGS) followed by bioinformatics tools has facilitated the identification of
9 numerous eukaryotic microorganisms in diverse hosts and environments (9). In particular, the
10 metatranscriptomics based on RNA-sequencing (RNA-seq) overcomes several obstacles
11 associated with virus identification (10,11). NGS techniques enable us to identify both known
12 and novel viruses in the host and provide several types of information, such as viral genomes
13 and mutations (12,13).

14 Recently, virus-associated studies using NGS have used a new term, “virome,” which
15 refers to a whole collection of viral nucleic acids in a specific host or a particular ecosystem
16 (14,15). In fact, virome studies provide several types of information associated with viral
17 communities, such as viral genomes, viral populations, mutations, phylogeny, evolution,
18 recombination, and interaction of viruses with the host or ecosystem, which have been widely
19 used in many such studies.

20 Viruses infecting fungi are referred to as mycoviruses (16). Most observed
21 mycoviruses consist of double-stranded (ds) RNA genomes; however, observed mycoviruses
22 with positive or negative single-stranded (ss) RNA and ssDNA genomes is gradually
23 increasing (17,18). As compared to other viruses infecting animals and plants, most known
24 mycoviruses lack an extracellular route for infection and movement proteins (19). Some of

1 identified mycoviruses infecting phytopathogenic fungi have the ability to reduce the fungal
2 hosts' pathogenicity, known as hypovirulence (16). Many novel mycoviruses are now being
3 identified based on metatranscriptomics, such as 66 novel mycoviruses with 15 distinct
4 lineages from five plant-pathogenic fungi (20) and 59 viruses from 44 different fungi (21).
5 Moreover, several studies have shown that fungal hosts are very often coinfecting by various
6 mycoviruses (21,22).

7 The rapid development and application of NGS techniques has increased the vast
8 amount of sequencing data in public databases, such as the Short Read Archive (SRA) at
9 NCBI, which are freely available. Many studies have demonstrated the usefulness of
10 transcriptomic data for virus identification and virome study without any *a priori* knowledge
11 of virus infection (21,23). Here, we carried out *in silico* bioinformatics analyses to identify
12 viruses infecting fungi using available fungal transcriptome data and addressed several
13 mycovirus-associated findings.

14

15 **MATERIALS AND METHODS**

16

17 **Download of fungal transcriptomes**

18 In order to identify virus-associated contigs (transcripts) in fungal transcriptomes, all
19 available fungal assembled contigs were downloaded from the TSA database
20 (<https://www.ncbi.nlm.nih.gov/genbank/tsa/>) using “fungi” as a query.

21

22 **Construction of viral genome and protein database**

23 All viral genomes and protein sequences were downloaded from NCBI's viral genome
24 website (<https://www.ncbi.nlm.nih.gov/genome/viruses/>) on October 10, 2018. The

downloaded sequences were used to set up a database for BLAST and DIAMOND using the following commands: “makeblastdb –in viral_genome.fasta –dbtype nucl –out viral_genome” and “diamond makedb –in viral_protein.faa -d viral_protein” (24).

Identification of virus-associated contigs

All fungal transcripts were subjected to BLASTX search against the viral protein database using the DIAMOND program with E-value 1e-10 as a cutoff (diamond blastx -d viral_protein -q fungal_contigs.fasta -e 1e-10 -k 1 -f 6 qseqid sseqid pident length mismatch gapopen qstart qend sstart send evalule bitscore qlen -o fungal_contigs_viral_protein.txt). Out of 3,733,874 fungal transcripts, 7,244 contigs showing similarity to viral proteins were extracted for further analysis. The 7,244 extracted contigs were subjected to BLASTX using DIAMOND with E-value 1e-10 as a cutoff against NCBI’s NR database (<ftp://ftp.ncbi.nlm.nih.gov/blast/db/FASTA/nr.gz>) to remove sequences derived from hosts and other organisms such as eukaryotes and bacteria. We ultimately obtained 468 virus-associated contigs.

Viral genome annotation

The 468 virus-associated contigs were again subjected to TBLASTX and BLASTN searches against the viral genome database to annotate the virus-associated contigs using E-value 1e-10. We selected virus-associated contigs larger than 1,000 bp for the prediction of open reading frames (ORFs) by the stand-alone version of ORFfinder in NCBI (<ftp://ftp.ncbi.nlm.nih.gov/genomes/TOOLS/ORFfinder/linux-i64/>). The predicted ORFs were compared to the related reference virus genome by BLASTP. Conserved domains in the virus-associated contigs were predicted by the SMART program (<http://smart.embl->

1 heidelberg.de/) (25).

2

3 **Phylogenetic tree construction**

4 We obtained 153 viral protein sequences composed of 107 viral replicase-associated
5 sequences, 17 CPs, 15 hypothetical proteins, six MPs, and 14 other proteins. We only used
6 replicase-associated protein sequences as representatives of identified viruses for the
7 phylogenetic analyses. To examine the phylogenetic relationship of identified viruses with
8 known viruses, 107 viral replicase proteins were divided into 16 viral families, one viral order,
9 and one unclassified group, resulting in 18 datasets. The viral replicase proteins in each
10 dataset were subjected to BLASTP search against the NR database in NCBI. We included all
11 matched protein sequences and removed redundant protein sequences.

12 Viral replicase sequences in each dataset were aligned using the L-INS-I method
13 implemented in MAFFT version 7.450 (26). We removed all ambiguously aligned sequences
14 by TrimAl (version 1.2) with the option “automated1” (27). We determined the best-fit model
15 of amino acid substitution in aligned sequences using IQ-TREE (28). The maximum
16 likelihood phylogenetic tree in each dataset was inferred using IQ-TREE with the ultrafast
17 bootstrap method (29) and the SH-aLRT branch test (30). The obtained phylogenetic trees
18 were visualized by FigTree (<http://tree.bio.ed.ac.uk/software/figtree/>).

19

20 **Estimation of viral abundance in fungal transcriptomes**

21 We downloaded all raw sequence data associated with the 43 fungal transcriptomes
22 containing virus-associated contigs from the SRA database in NCBI
23 (<https://www.ncbi.nlm.nih.gov/sra>). Unfortunately, some raw sequences were not available.

1 In many cases, there were several SRA datasets for a fungal transcriptome. We calculated
 2 viral abundance for an individual SRA dataset, not a single fungal transcriptome. The
 3 downloaded SRA data were converted into fastq files using the fastq-dump tool implemented
 4 in the SRA toolkit program (31). Each fastq file was mapped on the 468 virus-associated
 5 sequences using the Burrows-Wheeler Alignment (BWA) tool with the “BWA-MEM”
 6 algorithm (32). The mapped SAM file was subjected to pileup.sh in the BBMap program
 7 (<https://sourceforge.net/projects/bbmap/>) to extract the number of reads and FPKM values,
 8 which is a normalized estimation of gene expression considering the gene length and the
 9 sequencing depth in RNA-seq data. To calculate viral abundance in each fungal transcriptome,
 10 the number of virus-associated reads was divided by the total number of reads. We expected
 11 more reads from the large virus genomes and datasets with high sequencing depth. Therefore,
 12 we used FPKM values to calculate the proportion of viral abundance in a specific fungal
 13 transcriptome.

14 **Fungal transcriptome assembly and calculation of viral mutations**

16 For the virome study, we selected two previously studied fungal transcriptomes, *Monilinia*
 17 transcriptomes including 18 libraries and *Gigaspora margarita* transcriptomes including 21
 18 libraries. Both fungal transcriptomes were obtained from diverse fungal samples that were
 19 infected by many mycoviruses. However, the available assembled fungal transcriptomes were
 20 obtained by combining all libraries. Therefore, it was necessary to assemble an individual
 21 library for the virome study. Raw sequence reads from each condition were *de novo*
 22 assembled using the Trinity program (version 2.84) with default parameters (33). Then, each
 23 assembled transcriptome was subjected to BLASTN search against 468 virus-associated
 24 contigs. From each transcriptome, we extracted virus-associated contigs followed by

1 BLASTX search against the NR database to remove non-viral sequences. Finally, we
2 obtained virus-associated contigs from the individual condition. We mapped the raw
3 sequence reads against the virus-associated contigs in each condition to calculate viral
4 abundance and SNPs. Viral abundance in each condition was calculated by the number of
5 viral reads and FPKM values.

6 7 **Identification of virus mutations**

8 It is important to select proper reference sequences for the identification of virus mutations.
9 For example, if we use a known reference viral genome for mapping, we can find several
10 SNPs between two different viral genomes. However, the obtained SNPs are not viral
11 mutations. For that, we used viral sequences as references and raw sequence reads in the
12 identical condition for the calculation of SNPs. We obtained SNP results for 18 *Monilinia* and
13 21 *Gigaspora margarita* transcriptomes. SNPs were calculated as follows. For each condition,
14 raw sequence reads were mapped on the viral sequences obtained from the identical condition
15 using the BWA-MEM algorithm with default parameters. We converted the aligned reads in
16 the SAM file format into the BAM file format using SAMtools (ver. 1.3) (34). After sorting
17 and indexing the BAM file, SNP calling was performed using BCFtools (ver. 1.3) (35). We
18 obtained a VCF (Variant Call Format) file containing information on SNPs representing virus
19 mutations for each condition.

20 21 **RESULTS**

22 23 **Identification of virus-associated contigs from fungal transcriptomes**

24 To identify viruses from fungal transcriptomes, we obtained 3,733,874 fungal assembled

1 contigs (4,268,333,264 bp) from 126 different fungal transcriptomes in the Transcriptome
2 Shotgun Assembly (TSA) database (Fig. 1a and Table S1). Initially, all fungal contigs were
3 subjected to a BLASTX search against the viral protein database, resulting in 7,244 contigs
4 based on E-value $1e-10$ as a cutoff. After removing sequences from hosts and other organisms,
5 we obtained 468 virus-associated contigs (Fig. 1b and Table S2).

6 The BLAST results showed that the sequence identity (20.9–98.7%) and bit score
7 (36.6–7520.6) for the 468 virus-associated contigs against viral reference genomes were
8 relatively correlated (Fig. 1c). Most virus-associated contigs ranged from 201 to 1,000 bp
9 (227 contigs) followed by 2,001 to 3,000 bp (97 contigs) (Fig. 1d). Moreover, five contigs
10 were larger than 10 kb. Of the 126 fungal transcriptomes, only 43 included virus-associated
11 contigs (Fig. 1e and Table S1). The 43 fungal transcriptomes were assigned to five fungal
12 phyla, 18 orders, and 31 genera (Figs. 1f–1h). The most frequently identified fungal phyla
13 were *Basidiomycota* (16 transcriptomes), *Ascomycota* (13 transcriptomes), and
14 *Mucoromycota* (eight transcriptomes) (Fig. 1f). The top three fungal orders possessing
15 viruses were *Agaricales*, *Pucciniales*, and *Entomophthorales* (Fig. 1g). At the genus level,
16 *Monilinia*, *Podosphaera*, and *Entomophthora* were frequently infected by different viruses
17 (Fig. 1h).

18 A majority of virus-associated contigs were homologous to known viral proteins,
19 including RNA-dependent RNA polymerase (RdRp) (274 contigs) followed by polyprotein
20 (72 contigs) and hypothetical protein (34 contigs) (Fig. 1i and Table S2). Most virus-
21 associated contigs were derived from positive ssRNA genomes (214 contigs) and dsRNA
22 genomes (209 contigs) (Fig. 1j). Furthermore, some contigs were derived from viruses with
23 negative ssRNA, ssDNA, and dsDNA genomes.

24 The 468 virus-associated contigs were assigned to five viral orders, 21 families, 26

genera, and 88 species (Table S2). The frequently identified viral orders, families, and genera were *Tymovirales* (23 contigs) and *Picornavirales* (seven contigs); *Narnaviridae* (98 contigs), *Totiviridae* (95 contigs), and *Hypoviridae* (35 contigs) (Fig. 1k); and *Mitovirus* (84 contigs), *Hypovirus* (35 contigs), and *Totivirus* (16 contigs) (Fig. 1l), respectively.

Viral genome assembly and phylogeny of novel mycoviruses

We assembled 120 viral genomes divided into 16 viral families, one order, and unclassified viruses based on obtained virus-associated contigs (Table S3). Most viruses were novel viruses with low sequence identity with known viruses and had a single genome, except for four viruses with two RNA segments. Except for one ssDNA virus, all viruses had RNA genomes, which were further divided into unknown RNA (12 viruses), dsRNA (46 viruses), positive ssRNA (57 viruses), and negative ssRNA genomes (four viruses). Of the 18 identified virus genera, many viruses (28) were unclassified, while totiviruses (20) and ourmiaviruses (13) were two major viral genera.

We analyzed the genome organization of the 120 assembled viral genomes (Fig. 2a and Fig. S1). Most viruses encoded a single protein. For example, all viruses in the families *Botourmiaviridae*, *Narnaviridae*, *Partitiviridae*, *Bromoviridae*, *Circoviridae*, *Phasmaviridae*, and *Rhabdoviridae* encoded an RdRp. Out of nine viruses in the family *Gammaflexiviridae* encoding an RdRp, three viruses encoded RdRp and movement proteins (MPs). In the family *Totiviridae*, many identified viruses had both coat proteins (CPs) and RdRp proteins or a single CP or RdRp, except for the genome of *Uromyces totivirus D*, which encoded two hypothetical proteins. Some viruses in the families *Potyviridae*, *Fusariviridae*, *Secoviridae*, *Hypoviridae*, and *Tymoviridae* encoded a polyprotein. The genome organization for viruses in the family *Virgaviridae* and unclassified viruses was very diverse. For instance,

1 Entomophthora virgavirus A encoded a polyprotein and three hypothetical proteins, whereas
2 Leucocoprinus tobamovirus A encoded an RdRp and MP.

3 Based on 107 RdRp amino acid sequences, we generated a phylogenetic tree
4 revealing 16 viral families, one viral order, and one group with unclassified viruses (Fig. 2b
5 and Fig. S2). To reveal the phylogenetic relationships of the identified viruses in detail, we
6 generated 18 phylogenetic trees according to virus families (Figs. S3–S20). The six major
7 virus families identified were *Narnaviridae* (20 viruses), *Botourmiaviridae* (13 viruses),
8 *Totiviridae* (13 viruses), *Gammapflexiviridae* (nine viruses), *Hypoviridae* (seven viruses), and
9 *Partitiviridae* (seven viruses) (Fig. 2c). The phylogenetic trees revealed the genetic
10 relationships and host interactions of the identified viruses (Fig. 3c). For instance, all viruses
11 in the families *Botourmiaviridae*, *Gammapflexiviridae*, and *Hypoviridae* as well as nine
12 mitoviruses in the family *Narnaviridae* derived from this study were closely related to those
13 from fungal hosts. However, 11 narnaviruses, a totivirus, and two partitiviruses were grouped
14 with those from insects.

15 The 20 viruses identified in the family *Narnaviridae* were further divided into
16 narnaviruses (11 viruses) and mitoviruses (nine viruses) (Fig. S3). Most narnaviruses were
17 closely related to known narnaviruses from insects except for two *Puccinia* narnaviruses,
18 which were grouped with the *Fusarium poae* narnavirus 1. Interestingly, four narnaviruses
19 derived from trypanosomatids, which are parasites of insects, were clustered together with
20 those from insects and fungi. We identified 13 ourmiaviruses in the family *Botourmiaviridae*
21 from six fungal species in the same clade, which were distantly grouped with two
22 ourmiaviruses from plants (Fig. S4). Interestingly, four different *Monilinia* ourmiaviruses
23 were phylogenetically different from each other. Except for *Uromyces* totivirus D in the same
24 clade with two Wuhan insect viruses, 13 totiviruses from *Uromyces*, *Podosphaera*, and

1 *Phakopsora* genera were together with other totiviruses from fungi (Fig. S5). Of nine
2 gammaflexiviruses from *Leucocoprinus* species, *Leucocoprinus* gammaflexivirus E was
3 distantly related with other gammaflexiviruses (Fig. S6). Except for *Mycosphaerella*
4 hypovirus A, six hypoviruses were identified from *Monilinia* species (Fig. S7). The
5 phylogenetic tree showed that all known hypoviruses were identified from fungi except for
6 two hypoviruses from insects. Among the seven identified partitiviruses, *Entomophthora*
7 partitivirus D was closely related to other partitiviruses from insects (Fig. S8).

8 Two identified barnaviruses in this study were very distantly related with known
9 *Rhizoctonia solani* barnavirus 1 (Fig. S9). Out of six benyviruses, we identified two Hubei
10 Beny-like viruses from *Entomophthora muscae* causing a fungal disease in flies, while the
11 other four viruses were grouped with those from fungi (Fig. S10). We identified an anulavirus
12 in the family *Bromoviridae*, a circovirus in the family *Circoviridae*, and two fusariviruses
13 from fungi (Figs. S11–S13). The *Mucor* phasmavirus A in this study was grouped with
14 *Mucorales* RNA virus 1 and two other insect viruses (Fig. S14). Of two picornaviruses, the
15 novel *Pleurotus* picornavirus A from mushroom showed strong similarity to known
16 *Dicistroviridae* TZ-1 in the family *Dicistroviridae* from human blood (Fig. S15). *Uromyces*
17 potyvirus A and *Zymoseptoria* comovirus A were grouped with other viruses from plants (Fig.
18 S16 and Fig. S17). *Entomophthora* rhabdovirus A was closely related with other insect
19 rhabdoviruses (Fig. S18). Three tymoviruses from *Leucocoprinus* species and all six
20 unclassified viruses were grouped with those from fungi (Fig. S19 and Fig. S20).

21

22 **Viral abundance in different fungal transcriptomes**

23 We examined viral abundance in each fungal transcriptome based on FPKM (fragments per
24 kilobase of exon model per million reads mapped) values by mapping raw sequence reads

1 from SRA data on the 468 virus-associated contigs. Unfortunately, few SRA datasets were
2 available (Table S4). After calculating FPKM values for each SRA dataset, we found that the
3 SRA data from the FLX454 system were highly mapped on the virus-associated contigs due
4 to long read sizes (Table S5). Therefore, we excluded seven SRA datasets from further
5 analysis. Finally, we obtained FPKM values from 111 SRA datasets (Table S5). To simplify
6 the process, we combined the viral FPKM values for each identified virus species in the same
7 fungal species (Table S6). The 24 fungal species showed unique viral populations (Fig. 3a).
8 Mitoviruses were dominant viruses in transcriptomes of *Alternaria* species (AlM),
9 *Aureobasidium melanogenum* (AuM), *Cronartium ribicola* (CrR), *Eutiarosporella tritici-*
10 *australis* (EuT), and *Paraglomus brasilianum* (PaB). In some fungal species, a single virus
11 was identified, such as naravirus in *Cyathus bulleri* (CyB), phasmavirus in *Mucor*
12 *irregularis* (MuI), gemycircularvirus in *Pecoramyces ruminatum* (PeR), unclassified virus in
13 *Rhizopus oryzae* (RhO), and partitivirus in *Yarrowia* species (Ya3). *Leucoagaricus*
14 *gongylophorus* (LeG) contained several viruses; however, an unclassified virus was
15 dominantly present. *Racocetra castanea* (RaC) contained four different viruses, and of them,
16 endornavirus and hypovirus were dominantly present.

17 We compared the number of infecting viruses and the amount of viral abundance (Fig.
18 3b) and found that the number of infecting viruses was correlated with viral abundance. For
19 example, both GiM and LeG contained 24 different viruses with a high amount of viral
20 abundance. However, PeR was infected by a single virus, but its viral abundance was very
21 high. Next, we examined the frequency of identified viruses in different fungal species (Fig.
22 3c). Anulavirus, betapartitivirus, potexvirus, and potyvirus were identified in a single fungal
23 species, whereas mitovirus was identified in at least 17 fungal species. Moreover,
24 endornavirus (14 species), the unclassified virus (14 species), partitivirus (12 species),

barnavirus (11 species), and ourmiavirus (11 species) were frequently identified in different fungal species. Based on FPKM values, mitovirus showed the highest viral abundance followed by the unclassified virus and gemycircularvirus.

After combining virus-associated FPKM values, we examined the proportion of identified viruses within individual fungal species (Fig. 3d). The proportions of virus-associated FPKM values were very high in LeG (13%), PeR (12%), and MoFG (11%). Based on FPKM values, mitovirus (29%) was the most abundantly present virus followed by the unclassified virus (20%), gemycircularvirus (15%), ourmiavirus (8%), and hypovirus (8%) (Fig. 3e). Finally, we examined the proportion of mycoviruses in each fungal transcriptome based on sequence reads (Fig. 3f and Table S7). The proportion of virus-associated reads in most transcriptomes ranged from 0.01% to 0.1%; however, the proportion of virus-associated reads in two transcriptomes of *Leucoagaricus gongylophorus* was higher than 41% (Table S7).

Analyses of diverse *Monilinia* and *Gigaspora* viromes

We examined changes of viromes for two selected fungal transcriptome projects for *Monilinia* and *Gigaspora* species, respectively. *Monilinia* transcriptomes comprised 18 different samples established from three different hosts, four different regions, and three different *Monilinia* species: *M. fructicola*, *M. laxa*, and *M. fructigena* (36). The number of identified viruses ranged from one to 10 (Table S8). In particular, three samples from *M. fructigena* collected from pear in Emilia-Romagna were severely infected by many viruses (nine to 10 viruses) (Fig. 4a). In addition, the proportion of virus-associated reads in three *M. fructigena* transcriptomes was relatively high (2.3–6.1%) as compared to those of the other two *Monilinia* species (less than 1%).

1 We examined the proportion of infecting viruses in each *Monilinia* transcriptome
 2 (Fig. 4b). In general, the samples from the same host and region showed similar viral
 3 populations except for three *M. laxa* samples derived from cherry in Puglia showing very
 4 different viral populations. *M. fructicola* collected from cherry in two different regions
 5 showed region-specific viral populations. For example, mitovirus was dominant in *M.*
 6 *fructicola* from the Puglia region, while ourmiavirus was dominant in *M. fructicola* from the
 7 Campania region. Although different growth conditions and tissues were used, we did not
 8 find any possible effects on viral populations. Hypovirus was dominant in two samples from
 9 plum in the Basilicata region, whereas barnavirus was dominant in three samples from pear in
 10 the Emilia-Romagna region.

11 Next, we analyzed viromes of *Gigaspora margarita*, which comprised 21
 12 transcriptomes derived from 10 different conditions (Table S9) (37). *G. margarita* was
 13 infected by at least nine different viruses, and the proportion of virus-associated reads ranged
 14 from 0.06–0.33% (Fig. 4c). The GOU-13 sample was infected by only four viruses; however,
 15 the proportion of virus-associated reads was the highest (0.33%). Two frequently identified
 16 viruses in *G. margarita* were mitovirus and narnavirus (Fig. 4d). In particular, mitovirus was
 17 the most dominant virus in six out of 10 conditions. Although *G. margarita* was derived from
 18 a single strain, the viral populations in each condition were varied. Strigolactone is known to
 19 stimulate arbuscular mycorrhizal fungi by activating mitochondria (37). GOU-15, GDR-31,
 20 and GDR-34 samples were treated with strigolactone; however, we did not observe any
 21 significant changes in the viromes of *G. margarita* caused by strigolactone. This study was
 22 initially conducted to observe the symbiotic effects of an endobacterium to increase the
 23 bioenergetic potential of *G. margarita* (37). The number of infecting viruses and viral
 24 abundance were slightly reduced in three symbiotic conditions—GDR-16, GDR-36, and

1 GDR-38—as compared to the other seven pre-symbiotic conditions.

2 Viruses have high mutation rates as compared to their hosts. We examined virus
3 mutation frequency in two different viromes. Although *Monilinia* was infected by many
4 viruses, only three samples from *M. fructigena* collected from pear in Emilia-Romagna
5 showed many mutations (Table S10). In particular, barnavirus showed many mutations as
6 compared to other viruses (Fig. 4e). By contrast, several mutations were observed in most *G.*
7 *margarita* samples (Fig. 4f and Table S11). Interestingly, GOU15 and GDR-31 samples
8 treated with strigolactone had many mutations in several identified virus genomes. In GOU-
9 15 containing six viruses, barnavirus (43 mutations) had the highest number of mutations
10 followed by mitovirus (31 mutations) and parvovirus (22 mutations). In GDR-31 containing
11 four viruses having mutations, narnavirus (91 mutations) had the highest number of
12 mutations followed by partitivirus (35 mutations) and narnavirus (28 mutations). Interestingly,
13 the numbers of virus mutations for two symbiotic conditions, GOU-16 and GOU-38, were
14 zero and 10, respectively.

15

16 **DISCUSSION**

17 Here, we conducted a large-scale *in silico* mycovirus identification and mycovirome studies
18 using public fungal transcriptome datasets. Our study focused on mycoviruses, which have
19 not been well explored as compared to viruses infecting other eukaryotic kingdoms. Similarly,
20 a recent study also carried out *in silico* bioinformatics analyses using available fungal
21 transcriptomes identifying 59 mycoviral genomes (21). There were two main differences in
22 materials and methods between two studies. The first was the diversity of fungal
23 transcriptomes. In our study, we analyzed 126 fungal transcriptomes from at least 11 fungal
24 phyla whereas the recent study focused on Pezizomycotina fungi of the phylum *Ascomycota*.

1 The second was reference viral proteins for virus identification. To increase the number of
2 virus-associated contigs, we used all available viral proteins for the BLASTX search instead
3 of using viral RdRp sequences, revealing a wide range of mycoviruses encoding diverse viral
4 proteins. By contrast, the recent study only used six different viral RdRp families. Our
5 approach enabled us assemble 120 viral genomes and reveal mycoviruses classified into five
6 orders, 21 families, 26 genera, and 88 species, representing the widest range of mycoviral
7 taxonomy in a single study. Although both studies analyzed assembled mycoviral genomes
8 with phylogenetic trees and domain study, our study additionally showed viral abundance,
9 viral diversity, and mutation frequency using fungal transcriptomes. Our results might
10 provide a better picture of how much of the true overall mycovirome is visible within these
11 results.

12 To increase specificity for virus identification, BLASTX against the NR database
13 eliminated possible endogenous virus-like sequences, resulting in 468 virus-associated
14 contigs. Due to the low abundance of mycoviruses in fungal transcriptomes, the number of
15 mycovirus-associated reads was very low resulting that about half of the assembled virus-
16 associated contigs were less than 1 kb. Moreover, it is also possible that 83 of 126 fungal
17 transcriptomes that did not contain any virus-associated contigs might be infected by
18 unknown mycoviruses. The low abundance of mycoviruses in the fungal hosts could not be
19 identified by metatranscriptomics; however, they could be identified PCR based approaches
20 with virus-specific primers. Identification of virus by a single approach does not guarantee
21 virus infection in a living organism.

22 Most identified virus-associated contigs were derived from ssRNA and dsRNA
23 genomes; however, we identified viruses with negative ssRNA, ssDNA, and dsDNA genomes,
24 suggesting transcribed viral RNAs from DNA genomes by metatranscriptomics. For example,

the infection of gemycircularviruses and genomoviruses with a single circular DNA genome in fungi has been reported previously (18); however, this is the first report of identification of a mastrevirus and alphasatellite from fungal transcriptomes.

The phylogenetic tree revealed that many of the identified mycoviruses were originated from fungi; however, some mycoviruses had strong phylogenetic relationships with those from insects and plants. In particular, four known narnaviruses derived from trypanosomatids were closely related with those from insects and fungi, suggesting that the previously identified narnaviruses from insects might be derived from insect parasites (38). It is likely that the insect parasites were also included during RNA preparation from insect tissues.

Several virus families, including *Bromoviridae*, *Circoviridae*, and *Dicistroviridae*, have been identified from fungi for the first time. Interestingly, *Gigaspora* circovirus A with ssDNA genomes showed sequence similarity to only beak and feather disease virus, which causes a viral disease in birds (39). Circoviruses are circular ssDNAs that encode for two proteins bidirectional replication initiator protein (Rep) and CP. However, *Gigaspora* circovirus A has only a Rep. Furthermore, *Gigaspora* circovirus A was identified in all 10 *Gigaspora* transcriptomes. This result confirmed the possible infection of a circovirus in the fungus. The strong similarity between the novel *Pleurotus* picornavirus A and known *Dicistroviridae* TZ-1 from the blood of febrile Tanzanian children was somewhat interesting (40). However, other dicistroviruses from insects were grouped together with picornaviruses from insects, suggesting the presence of dicistrovirus in the mushroom. In addition, parvoviruses and flaviviruses, which infect both vertebrate and invertebrate hosts, were detected in fungi for the first time.

Uromyces potyvirus A and *Zymoseptoria* comovirus A showed strong similarity to

1 other viruses from plants, indicating that both fungi transmit plant viruses that infect both
2 fungi and plants. Fungal transmission of a plant virus has been reported (41). Moreover,
3 plants are known hosts for viruses in the family *Virgaviridae*, which were also identified in
4 fungi, suggesting that fungi could be alternative hosts. A previous study reporting the
5 replication of endophytic mycovirus in plant cells supports our finding (42).

6 The genome organization of the 120 mycoviromes indicates that most mycoviruses
7 had simple genome structures. For instance, many of the identified mycoviruses had an RdRp
8 required for viral replication. In addition, we found that the genome structures of some
9 viruses in the families *Barnaviridae*, *Benyviridae*, *Gammapflexiviridae*, *Virgaviridae*, and
10 *Tymoviridae* were not identical, indicating changes of viral genome structures during the
11 mycovirus's evolution and host adaptation.

12 A few recent studies revealed the complete genome of a large number of viruses from
13 vertebrates and invertebrates using transcriptomic and complementary approaches (3,4). By
14 contrast, in our *in silico* mycovirome study, we faced several obstacles to obtaining complete
15 mycoviral genomes from only transcriptome data because of the low titer of mycoviruses in
16 the fungi and unavailability of original samples. However, 468 mycovirus-associated
17 sequences, in which 120 viral genomes were assembled, provided a comprehensive overview
18 of the distribution of mycovirus taxonomy and their hosts. It was also surprising that the
19 fungi could be the hosts of numerous viruses, as shown in GiM and LeG, which were infected
20 by at least 24 different viruses. In addition, individual mycoviromes showed at least 10 and
21 eight different viruses infecting *Monilinia* and *Gigaspora* species, respectively, revealing the
22 coinfection of mycoviruses in many fungal hosts.

23 The virome study based on the mapping of raw sequence reads revealed a possible
24 correlation between the number of infecting viruses and viral abundance. This result suggests

1 that the competition of diverse coinfecting mycoviruses promoted the viral replication.
 2 Although the coinfection of diverse mycoviruses in the fungi was common in our study, most
 3 viromes had a dominant virus species. Our study identified the most common mycoviruses,
 4 such as mitovirus, endornavirus, benyvirus, ourmiavirus, and partitivirus, with a wide range
 5 of fungal hosts, as described previously (20,21). However, several viruses, including
 6 anulavirus, betapartitivirus, potexvirus, and potyvirus, had a limited range of hosts.

7 The single nucleotide polymorphism (SNP) analysis showed that there were
 8 differences in virus mutation frequency among coinfecting mycoviruses, and most
 9 mycoviruses did not show strong mutation except for a few mycoviruses, such as barnavirus,
 10 mitovirus, and narnavirus.

11 The study of *Monilinia* viromes revealed that there were strong deviations in the
 12 composition and viral abundance of viruses that originated from identical samples. This
 13 might have been caused by virus abundance, RNA extraction, library preparation, or RNA-
 14 seq. In addition, hosts and sample regions played an important role in viral populations in
 15 different *Monilinia* species. Although different dark and light conditions and tissues were
 16 used for sample preparation, it was not sufficient to determine their effects on the virome
 17 without biological replicates.

18 The study of *Gigaspora* viromes revealed that the chemical strigolactone might
 19 promote virus replication and mutations, while symbiosis with endobacteria might suppress
 20 virus replication and mutations. Strigolactones, plant hormones extracted from plant roots,
 21 play an important role in the symbiotic interaction between plants and arbuscular mycorrhizal
 22 fungi. However, our results demonstrated negative effects of strigolactones on fungi and
 23 mycoviruses. The strong symbiotic interaction between endobacteria and *Gigaspora* species
 24 inhibited mycovirus replication. Thus, the effects of strigolactones and the symbiosis of

1 endobacteria on mycovirus lifecycles should be further elucidated.

2 As shown in our study, the compositions and populations of mycoviruses were more
3 complex than we expected. This study is the largest comprehensive mycovirome study
4 shedding light on an array of mycovirome-associated topics, including virus identification,
5 fungal host ranges, taxonomy of mycoviruses, viral genome assembly, phylogenetic analyses,
6 viral abundance, mutations, and changes of mycoviral populations in different conditions.
7 Moreover, this study provides many clues and information to reveal the diversity and host
8 distribution of mycoviruses.

10 **DATA AVAILABILITY**

11 Viral genome sequences for 120 mycoviruses were deposited in GenBank with the following
12 accession numbers: MK231015–MK231133 and MK940812.

14 **SUPPLEMENTARY DATA**

15 **Supplementary Figures**

16 Figure S1. Genome organization for identified viruses in families *Barnaviridae*, *Benyviridae*,
17 *Bromoviridae*, *Circoviridae*, *Fusariviridae*, *Hypoviridae*, *Phasmaviridae*, *Potyviridae*,
18 *Rhabdoviridae*, and *Secoviridae* and unclassified viruses.

19 Figure S2. Phylogenetic tree of 107 identified viruses using RdRp sequences.

20 Figure S3. Phylogenetic tree of 20 viruses in family *Narnaviridae* and related viruses.

21 Figure S4. Phylogenetic tree of 13 viruses in family *Botourmiaviridae* and related viruses.

22 Figure S5. Phylogenetic tree of 14 viruses in family *Totiviridae* and related viruses.

23 Figure S6. Phylogenetic tree of nine viruses in family *Gammaplexiviridae* and related viruses.

24 Figure S7. Phylogenetic tree of seven viruses in family *Hypoviridae* and related viruses.

- 1 Figure S8. Phylogenetic tree of seven viruses in family *Partitiviridae* and related viruses.
- 2 Figure S9. Phylogenetic tree of two viruses in family *Barnaviridae* and related viruses.
- 3 Figure S10. Phylogenetic tree of six viruses in family *Benyviridae* and related viruses.
- 4 Figure S11. Phylogenetic tree of a virus in family *Bromoviridae* and related viruses.
- 5 Figure S12. Phylogenetic tree of a virus in family *Circoviridae* and related viruses.
- 6 Figure S13. Phylogenetic tree of two viruses in family *Fusariviridae* and related viruses.
- 7 Figure S14. Phylogenetic tree of a virus in the family *Phasmaviridae* and related viruses.
- 8 Figure S15. Phylogenetic tree of two viruses in the order *Picornavirales* and related viruses.
- 9 Figure S16. Phylogenetic tree of a virus in the family *Potyviridae* and related viruses.
- 10 Figure S17. Phylogenetic tree of a virus in the family *Secoviridae* and related viruses.
- 11 Figure S18. Phylogenetic tree of a virus in the family *Rhabdoviridae* and related viruses.
- 12 Figure S19. Phylogenetic tree of three viruses in the family *Tymoviridae* and related viruses.
- 13 Figure S20. Phylogenetic tree of six unclassified viruses and related viruses.
- 14 Supplementary Tables
- 15
- 16 Table S1. Summary of fungal transcriptomes used for virus identification.
- 17 Table S2. Summary of BLASTX results for 468 virus-associated contigs against viral
- 18 proteins.
- 19 Table S3. Detailed information of 120 viral genomes assembled from fungal transcriptomes
- 20 in this study.
- 21 Table S4. Information of SRA data used for virome study.
- 22 Table S5. Mapping results of raw sequence reads for each SRA dataset on 468 virus-
- 23 associated contigs.
- 24 Table S6. FPKM values for identified viruses in each fungal species.

1 Table S7. Mapping results of raw sequence reads for each SRA dataset on 468 virus-
2 associated contigs.

3 Table S8. Detailed information for *Monilinia* viromes.

4 Table S9. Detailed information for *Gigaspora margarita* viromes.

5 Table S10. Number of identified mutations for viruses identified from *Monilinia*
6 transcriptomes.

7 Table S11. Number of identified mutations for viruses identified from *Monilinia*
8 transcriptomes.

9

10 **FUNDING**

11 This work was supported by the National Research Foundation of Korea (NRF) grant funded
12 by the Korea government (MSIT) (No. NRF-2018R1D1A1B07043597) and the support of
13 "Cooperative Research Program for Agriculture Science and Technology Development
14 (Project No. PJ01498301)" Rural Development Administration, Republic of Korea.

15

16 **CONFLICT OF INTEREST**

17 The authors declare that they have no competing interests.

18

19 **ACKNOWLEDGEMENTS**

20 I (Won Kyong Cho) would like to express our deepest gratitude to Hyang Sook Kim, Mi
21 Kyong Kim, and Jin Kyong Cho for all their support with the current project. This work is
22 dedicated to the memory of my father, Tae Jin Cho (1946–2015).

23

24 **REFERENCES**

- 1
- 2 1. Lefkowitz, E.J., Dempsey, D.M., Hendrickson, R.C., Orton, R.J., Siddell, S.G. and
- 3 Smith, D.B. (2017) Virus taxonomy: the database of the International Committee on
- 4 Taxonomy of Viruses (ICTV). *Nucl Acid Res*, **46**, D708-D717.
- 5 2. Monier, A., Claverie, J.-M. and Ogata, H. (2008) Taxonomic distribution of large
- 6 DNA viruses in the sea. *Genome Biol*, **9**, R106.
- 7 3. Shi, M., Lin, X.-D., Tian, J.-H., Chen, L.-J., Chen, X., Li, C.-X., Qin, X.-C., Li, J.,
- 8 Cao, J.-P. and Eden, J.-S. (2016) Redefining the invertebrate RNA virosphere. *Nature*,
- 9 **540**, 539.
- 10 4. Shi, M., Lin, X.-D., Chen, X., Tian, J.-H., Chen, L.-J., Li, K., Wang, W., Eden, J.-S.,
- 11 Shen, J.-J. and Liu, L. (2018) The evolutionary history of vertebrate RNA viruses.
- 12 *Nature*, **556**, 197.
- 13 5. Mushegian, A., Shipunov, A. and Elena, S.F. (2016) Changes in the composition of
- 14 the RNA virome mark evolutionary transitions in green plants. *BMC Biol*, **14**, 68.
- 15 6. Shkoporov, A.N. and Hill, C. (2019) Bacteriophages of the human gut: the “known
- 16 unknown” of the microbiome. *Cell Host Microbe*, **25**, 195-209.
- 17 7. Prussin, A.J., Marr, L.C. and Bibby, K.J. (2014) Challenges of studying viral aerosol
- 18 metagenomics and communities in comparison with bacterial and fungal aerosols.
- 19 *FEMS Microbiol Lett*, **357**, 1-9.
- 20 8. Kumar, N., Barua, S., Riyesh, T., Chaubey, K.K., Rawat, K.D., Khandelwal, N.,
- 21 Mishra, A.K., Sharma, N., Chandel, S.S. and Sharma, S. (2016) Complexities in
- 22 isolation and purification of multiple viruses from mixed viral infections: viral
- 23 interference, persistence and exclusion. *PLoS One*, **11**, e0156110.
- 24 9. Nowrousian, M. (2010) Next-generation sequencing techniques for eukaryotic

- 1 microorganisms: sequencing-based solutions to biological problems. *Eukaryotic Cell*,
2 **9**, 1300-1310.
- 3 10. Zhang, Y.-Z., Shi, M. and Holmes, E.C. (2018) Using metagenomics to characterize
4 an expanding virosphere. *Cell*, **172**, 1168-1172.
- 5 11. Ho, T. and Tzanetakis, I.E. (2014) Development of a virus detection and discovery
6 pipeline using next generation sequencing. *Virology*, **471**, 54-60.
- 7 12. Yamashita, A., Sekizuka, T. and Kuroda, M. (2016) VirusTAP: viral genome-targeted
8 assembly pipeline. *Front Microbiol*, **7**, 32.
- 9 13. Prosperi, M.C., Prosperi, L., Bruselles, A., Abbate, I., Rozera, G., Vincenti, D.,
10 Solmone, M.C., Capobianchi, M.R. and Ulivi, G. (2011) Combinatorial analysis and
11 algorithms for quasispecies reconstruction using next-generation sequencing. *BMC*
12 *Bioinform*, **12**, 5.
- 13 14. Watson, S.J., Welkers, M.R., Depledge, D.P., Coulter, E., Breuer, J.M., de Jong, M.D.
14 and Kellam, P. (2013) Viral population analysis and minority-variant detection using
15 short read next-generation sequencing. *Philos Trans Royal Soc B*, **368**, 20120205.
- 16 15. Paez-Espino, D., Eloie-Fadrosh, E.A., Pavlopoulos, G.A., Thomas, A.D., Huntemann,
17 M., Mikhailova, N., Rubin, E., Ivanova, N.N. and Kyrpides, N.C. (2016) Uncovering
18 Earth's virome. *Nature*, **536**, 425.
- 19 16. Nuss, D.L. (2005) Hypovirulence: mycoviruses at the fungal-plant interface. *Nat Rev*
20 *Microbiol*, **3**, 632.
- 21 17. Liu, L., Xie, J., Cheng, J., Fu, Y., Li, G., Yi, X. and Jiang, D. (2014) Fungal negative-
22 stranded RNA virus that is related to bornaviruses and nyaviruses. *Proc Nat Acad Sci*,
23 **111**, 12205-12210.
- 24 18. Yu, X., Li, B., Fu, Y., Jiang, D., Ghabrial, S.A., Li, G., Peng, Y., Xie, J., Cheng, J. and

- 1 Huang, J. (2010) A geminivirus-related DNA mycovirus that confers hypovirulence to
2 a plant pathogenic fungus. *Proc Nat Acad Sci*, **107**, 8387-8392.
- 3 19. Son, M., Yu, J. and Kim, K.-H. (2015) Five questions about mycoviruses. *PLoS*
4 *Pathog*, **11**, e1005172.
- 5 20. Marzano, S.-Y.L., Nelson, B.D., Ajayi-Oyetunde, O., Bradley, C.A., Hughes, T.J.,
6 Hartman, G.L., Eastburn, D.M. and Domier, L.L. (2016) Identification of diverse
7 mycoviruses through metatranscriptomics characterization of the viromes of five
8 major fungal plant pathogens. *J Virol*, **90**, 6846-6863.
- 9 21. Gilbert, K.B., Holcomb, E.E., Allscheid, R.L. and Carrington, J.C. (2019) Hiding in
10 plain sight: New virus genomes discovered via a systematic analysis of fungal public
11 transcriptomes. *PLoS One*, **14**, e0219207.
- 12 22. Thapa, V. and Roossinck, M.J. (2019) Determinants of Coinfection in the
13 Mycoviruses. *Frontiers in cellular and infection microbiology*, **9**, 169.
- 14 23. Jo, Y., Choi, H., Cho, J.K., Yoon, J.-Y., Choi, S.-K. and Cho, W.K. (2015) In silico
15 approach to reveal viral populations in grapevine cultivar Tannat using transcriptome
16 data. *Sci Rep*, **5**, 15841.
- 17 24. Buchfink, B., Xie, C. and Huson, D.H. (2015) Fast and sensitive protein alignment
18 using DIAMOND. *Nat Methods*, **12**, 59.
- 19 25. Letunic, I., Doerks, T. and Bork, P. (2011) SMART 7: recent updates to the protein
20 domain annotation resource. *Nucl Acids Res*, **40**, D302-D305.
- 21 26. Katoh, K. and Standley, D.M. (2013) MAFFT multiple sequence alignment software
22 version 7: improvements in performance and usability. *Mol Biol Evol*, **30**, 772-780.
- 23 27. Capella-Gutiérrez, S., Silla-Martínez, J.M. and Gabaldón, T. (2009) trimAl: a tool for
24 automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics*,

- 1 **25**, 1972-1973.
- 2 28. Nguyen, L.-T., Schmidt, H.A., von Haeseler, A. and Minh, B.Q. (2014) IQ-TREE: a
3 fast and effective stochastic algorithm for estimating maximum-likelihood
4 phylogenies. *Mol Biol Evol*, **32**, 268-274.
- 5 29. Hoang, D.T., Chernomor, O., Von Haeseler, A., Minh, B.Q. and Vinh, L.S. (2017)
6 UFBoot2: improving the ultrafast bootstrap approximation. *Mol Biol Evol*, **35**, 518-
7 522.
- 8 30. Guindon, S., Dufayard, J.-F., Lefort, V., Anisimova, M., Hordijk, W. and Gascuel, O.
9 (2010) New algorithms and methods to estimate maximum-likelihood phylogenies:
10 assessing the performance of PhyML 3.0. *Sys Biol*, **59**, 307-321.
- 11 31. Kodama, Y., Shumway, M. and Leinonen, R. (2011) The Sequence Read Archive:
12 explosive growth of sequencing data. *Nucl Acids Res*, **40**, D54-D56.
- 13 32. Li, H. and Durbin, R. (2010) Fast and accurate long-read alignment with Burrows-
14 Wheeler transform. *Bioinformatics*, **26**, 589-595.
- 15 33. Grabherr, M.G., Haas, B.J., Yassour, M., Levin, J.Z., Thompson, D.A., Amit, I.,
16 Adiconis, X., Fan, L., Raychowdhury, R. and Zeng, Q. (2011) Full-length
17 transcriptome assembly from RNA-Seq data without a reference genome. *Nat Biotech*,
18 **29**, 644.
- 19 34. Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G.,
20 Abecasis, G. and Durbin, R. (2009) The sequence alignment/map format and
21 SAMtools. *Bioinformatics*, **25**, 2078-2079.
- 22 35. Li, H. (2011) A statistical framework for SNP calling, mutation discovery, association
23 mapping and population genetical parameter estimation from sequencing data.
24 *Bioinformatics*, **27**, 2987-2993.

- 1 36. Angelini, R.M.D.M., Abate, D., Rotolo, C., Gerin, D., Pollastro, S. and Faretra, F.
2 (2018) De novo assembly and comparative transcriptome analysis of *Monilinia*
3 *fruticola*, *Monilinia laxa* and *Monilinia fructigena*, the causal agents of brown rot on
4 stone fruits. *BMC Genomics*, **19**, 436.
- 5 37. Salvioli, A., Ghignone, S., Novero, M., Navazio, L., Bagnaresi, P. and Bonfante, P.
6 (2016) Symbiosis with an endobacterium increases the fitness of a mycorrhizal fungus,
7 raising its bioenergetic potential. *ISME J*, **10**, 130.
- 8 38. Grybchuk, D., Kostygov, A.Y., Macedo, D.H., Votýpka, J., Lukeš, J. and Yurchenko, V.
9 (2018) RNA Viruses in *Blechnomonas* (Trypanosomatidae) and Evolution of
10 *Leishmaniavirus*. *mBio*, **9**, e01932-01918.
- 11 39. Niagro, F., Forsthoefel, A., Lawther, R., Kamalanathan, L., Ritchie, B., Latimer, K.
12 and Lukert, P. (1998) Beak and feather disease virus and porcine circovirus genomes:
13 intermediates between the geminiviruses and plant circoviruses. *Arch Virol*, **143**,
14 1723-1744.
- 15 40. Cordey, S., Laubscher, F., Hartley, M.-A., Junier, T., Pérez-Rodriguez, F.J., Keitel, K.,
16 Vieille, G., Samaka, J., Mlaganile, T. and Kagoro, F. (2019) Detection of
17 dicistroviruses RNA in blood of febrile Tanzanian children. *Emerg Microbes Infect*, **8**,
18 613-623.
- 19 41. Andika, I.B., Wei, S., Cao, C., Salaipeth, L., Kondo, H. and Sun, L. (2017)
20 Phytopathogenic fungus hosts a plant virus: A naturally occurring cross-kingdom viral
21 infection. *Proc Nat Acad Sci*, **114**, 12267-12272.
- 22 42. Nerva, L., Varese, G., Falk, B. and Turina, M. (2017) Mycoviruses of an endophytic
23 fungus can replicate in plant cells: Evolutionary implications. *Sci Rep*, **7**, 1908.

1 **Figures**

2

3 **Figure 1.** Identification of virus-associated contigs from fungal transcriptomes.

4 (a) Schematic representation of bioinformatics workflow to identify virus-associated contigs
5 from fungal transcriptomes. (b) Proportion of identified virus-associated contigs in each
6 major step. (c) Distribution of protein identity (left) and bit score (right) for 468 virus-

1 associated contigs. (d) Size distribution of 468 identified virus-associated contigs. (e)
2 Number of fungal transcriptomes with and without viruses, respectively. Number of fungal
3 transcriptomes containing viruses according to fungal phylum (f), order (g), and genus (h).
4 Classification of virus-associated contigs according to viral proteins (i) and virus genome
5 type (j). Classification of 468 virus-associated contigs according to viral family (k) and genus
6 (l).
7

Figure 2. Phylogenetic trees and genome organization for assembled viruses.

(a) Genome organization for identified viruses in seven viral families and one viral order. Each virus protein was visualized by different colored boxes. The number on the left in each virus corresponds to the number in Table S2. The number on the right indicates virus genome size. Genome organization for viruses in other families can be found in Fig. S2. (b) Phylogenetic tree using RdRp sequences of 107 identified viruses. Amino acid sequences were aligned with MAFFT and used for phylogenetic tree construction using maximum

1 likelihood method and the LG+I+G4 protein substitution model implemented in IQ-TREE.
2 The phylogenetic tree is midpoint rooted using FigTree. Viruses were indicated by different
3 colors according to corresponding virus family. The detailed phylogenetic tree can be found
4 in Fig. S2. (c) Phylogenetic trees of identified viruses belonging to six major families and
5 related viruses. RdRp sequences of identified viruses and related viruses were aligned by
6 MAFFT followed by sequence trimming using TrimAL. Maximum likelihood phylogenetic
7 trees for individual virus families were constructed using IQ-TREE. The protein substitution
8 model was indicated. Red lines indicate viruses identified from our study. Viruses from plants,
9 fungi, and insects were indicated by different colored boxes. The detailed phylogenetic trees
10 can be found in Figs. S3–S8.

11

12

Figure 3. Number of identified viruses and viral abundance in individual fungal species.

(a) Proportion of identified viruses based on viral abundance. Viral abundance of each virus in identified fungal transcriptomes was calculated by FPKM values. FPKM values for identified viruses in same fungal species were combined for simplicity. (b) Number of identified viruses and total viral abundance in each fungal species. (c) Number of viruses infecting fungal transcriptomes for individual virus species and total viral abundance of

- 1 individual virus species. (d) Proportion of fungal species according to viral abundance. (e)
- 2 Proportion of identified viruses according to viral abundance. (f) Number of fungal
- 3 transcriptomes according to viral abundance. Viral abundance in each fungal transcriptome
- 4 was measured by number of virus-associated reads divided by total reads.

Figure 4. Analyses of viromes for three different *Monilinia* species and *Gigaspora margarita*.

(a) Number of identified viruses (left) and proportion of viral abundance (%) (right) in 18 different *Monilinia* transcriptomes. (b) Proportion of identified viruses in each *Monilinia* transcriptome. (c) Number of identified viruses (left) and proportion of viral abundance (%)

1 (right) in 10 different *G. margarita* transcriptomes. (d) Proportion of identified viruses in
2 each *G. margarita* transcriptome. Number of identified virus mutations for individual virus
3 species in each *Monilinia* (e) and *G. margarita* (f) transcriptome.

4