

RESEARCH ARTICLE

Hiding in plain sight: New virus genomes discovered via a systematic analysis of fungal public transcriptomes

Kerrigan B. Gilbert, Emily E. Holcomb, Robyn L. Allscheid¹, James C. Carrington¹*

Donald Danforth Plant Science Center, Saint Louis, Missouri, United States of America

¹ Current address: National Corn Growers Association, Chesterfield, Missouri, United States of America

* jcarrington@danforthcenter.org



OPEN ACCESS

Citation: Gilbert KB, Holcomb EE, Allscheid RL, Carrington JC (2019) Hiding in plain sight: New virus genomes discovered via a systematic analysis of fungal public transcriptomes. PLoS ONE 14(7): e0219207. <https://doi.org/10.1371/journal.pone.0219207>

Editor: Maureen J. Donlin, Saint Louis University, UNITED STATES

Received: May 10, 2019

Accepted: June 18, 2019

Published: July 24, 2019

Copyright: © 2019 Gilbert et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: Viral genomic nucleotide sequences are available from the GenBank database (accession numbers: MK279429-MK279511).

Funding: This work was supported by the National Institute of Health (NIH) Institute, National Institute of Allergy and Infectious Diseases (NIAID), AI043288 (<http://www.nih.gov/>) to JCC. The funder had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Abstract

The distribution and diversity of RNA viruses in fungi is incompletely understood due to the often cryptic nature of mycoviral infections and the focused study of primarily pathogenic and/or economically important fungi. As most viruses that are known to infect fungi possess either single-stranded or double-stranded RNA genomes, transcriptomic data provides the opportunity to query for viruses in diverse fungal samples without any *a priori* knowledge of virus infection. Here we describe a systematic survey of all transcriptomic datasets from fungi belonging to the subphylum Pezizomycotina. Using a simple but effective computational pipeline that uses reads discarded during normal RNA-seq analyses, followed by identification of a viral RNA-dependent RNA polymerase (RdRP) motif in *de novo* assembled contigs, 59 viruses from 44 different fungi were identified. Among the viruses identified, 88% were determined to be new species and 68% are, to our knowledge, the first virus described from the fungal species. Comprehensive analyses of both nucleotide and inferred protein sequences characterize the phylogenetic relationships between these viruses and the known set of mycoviral sequences and support the classification of up to four new families and two new genera. Thus the results provide a deeper understanding of the scope of mycoviral diversity while also increasing the distribution of fungal hosts. Further, this study demonstrates the suitability of analyzing RNA-seq data to facilitate rapid discovery of new viruses.

Introduction

Advances in low-cost, high-throughput sequencing technologies has revolutionized our ability to query the molecular world. Transcriptome projects are of particular use by providing targeted information about gene expression and improving the annotation of genome projects. Indeed, a number of “1K” transcriptome projects are underway, including the 1000 Plant Genome project, Fish-T1K project, 1K Insect Transcriptome Evolution (1KITE) project, and 1000 Fungal Genome and Transcriptome project. These large scale, collaborative projects have contributed to the vast amount of sequencing data available. As of February 2018, more than

Competing interests: The authors have declared that no competing interests exist.

16,000 petabases have been deposited in the Short Read Archive (SRA) at NCBI, with over 6,000 petabases available as open-access data.

Beyond revealing gene expression patterns within the target organism, these datasets are a source for additional insights as novel RNA sequences, specifically viral genomic sequences, have been identified within them. In an early example, analysis of RNA-seq data from a soybean cyst nematode revealed the presence of four different RNA viruses [1]. A more broad scale analysis of 66 non-angiosperm RNA-seq datasets, produced as part of the 1000 Plant Genome project, identified viral sequences in 28 plant species [2]. Additionally, the first viral sequences for any member of the family of spiders *Nephilidae* were identified through a secondary analysis of the RNA-seq data produced during the spider genome sequencing project [3].

While over 200 mycoviruses have been described, the full depth of the mycoviral landscape remains to be determined as certain biases have impacted the discovery of mycoviral sequences. Commonly infections of fungi appear cryptic, having little to no apparent impact on fungal health, thus occluding the presence of the virus. Additionally, many of the fungi studied are important human or plant pathogens, but this represents only a fraction of the known diversity of fungi. Advancements in mycovirus discoveries have been aided by screening the transcriptomes of various collections of fungi; study systems include those fungi associated with the seagrass *Posidonia oceanica* [4], five common plant-pathogenic fungal species [5], and soybean phyllosphere phytobiomes [6].

In the biosphere, fungi are not found in isolation; instead, they are members of complex environments that include members from some or all other kingdoms of life. Characterizing the responses and contributions that fungi make to these environments also includes a full understanding of the dynamics at play within the fungi. We hypothesized that mycoviral infections are more pervasive than currently known, and as such, undertook a systematic approach to identify new mycoviral species by querying the publicly available fungal transcriptomic datasets at the Short Read Archive hosted at NCBI. Our bioinformatic approach provides a robust method for the identification of new viral species from raw sequencing data as we identified 59 viral genomes, 52 of which are described here for the first time.

Results and discussion

Viral RNA-dependent RNA polymerases (RdRP) are essential enzymes in the genomes of RNA viruses that have no DNA stage. The domain organization of this protein is described as a right hand, where three subdomains are the palm, fingers, and thumb [7]. While some of the motifs found within these subdomains are conserved among all polymerase enzymes, specific structures and motifs are found only in RNA-dependent RNA polymerases. In addition to unique signatures within the sequences themselves, virus-like RdRPs are not encoded in the DNA genomes of eukaryotic organisms, thus viral RdRPs can be specifically targeted as a marker for sequence-based analyses.

We previously determined that contigs assembled from fungal RNA-seq data can be queried for viral RdRP-specific domains [8]. As such, we undertook a systematic analysis of all RNA-seq datasets from Pezizomycotina fungi (division: Ascomycota; subkingdom: Dikarya) available at the Short Read Archive (SRA) at the National Center for Biotechnology Information (NCBI) on July 21, 2017. While the subphylum Saccharomycotina contained a greater number of sequencing datasets, the subphylum Pezizomycotina was chosen for two reasons: it is host to the second highest number of RNA-seq data and it contained a more diverse set of fungal hosts since over 85% of Saccharomycotina datasets are from strains of *Saccharomyces*

cerevisiae. Datasets from the remaining Ascomycota subphylums, the phylum Basidiomycota, and unclassified fungi remain to be analyzed.

Querying for datasets from within the Pezizomycotina sub-phylum resulted in an initial list of 1,127 unique BioProjects, which was manually curated to identify 569 BioProjects suitable for this analysis (S1 Table). Samples were excluded for a variety of reasons including: individually listed BioProjects were determined to belong to larger groups of experiments using the same fungal strain and were therefore combined and a single representative sample was selected; samples containing a mix RNA from fungi and other kingdoms were excluded as the host of any putative viral sequences could not be determined; samples where the read sequences were too short for assembly or were unavailable for download.

Virus-like RdRP discovery pipeline

A computational pipeline was created that automated the steps from SRA data download through to identification of RdRP-containing contig sequences. The majority of the steps used publicly available programs in conjunction with a small number of custom Perl scripts. The overall structure of the pipeline was optimized to efficiently work within a HTCondor-managed system including the use of the DAGMan (Directed Acyclic Graph Manager) meta-scheduler to coordinate the required dependencies between pipeline steps.

Due to the highly specific nature of viral RdRP protein sequences, the key feature of this pipeline was the use of a customized Pfam database which included the viral RdRP families: RdRP_1, RdRP_2, RdRP_3, RdRP_4, RdRP_5, and Mitovir_RNA_pol. To err on the side of over-inclusiveness, HMMER *hmmsearch* was run with an e-value cut-off of 10.0, allowing for the identification of any sequence with limited similarity to the viral protein domains. Putative RdRP sequence(s) were identified in a total of 1,067 contigs from 284 of the 569 samples. The actual number of viruses was expected to be less than this total number, either because of false positive sequences or because multiple contigs originated from the same the viral genome. To confirm similarity to known viral sequences and rule out false positives, contigs were evaluated via a BLASTX search versus the 'nr' database at NCBI.

Summary of identified viral sequences

In total, 59 complete, RNA mycoviral genomes were identified in 47 of the 569 SRA samples analyzed. Across these datasets, there is an uneven distribution of data being generated within the Pezizomycotina subphylum whereby certain Classes are more represented than others; this is illustrated by the height of the band within the "Fungal Class" column in Fig 1. For instance, fungi within the Sordariomycetes and Eurotiomycetes represent 73% of the sequencing projects while the remaining 23% of datasets are spread across the remaining six Classes. Unsurprisingly, fungi known to be key pathogens of plants and humans are present within the Sordariomycetes and Eurotiomycetes, including the well-studied genera *Fusarium*, *Aspergillus*, and *Penicillium*. Indeed, seven genera alone account for 51% of all sequencing projects analyzed in this study, with the remaining 49% of datasets spread among 176 other genera. Similar patterns are also apparent among the set of known mycoviruses isolated from fungi within the Pezizomycotina (Virus-HostDB, Version 91). For instance, fungi from the Sordariomycetes and Eurotiomycetes classes host over half of the known Pezizomycotina mycoviruses. Conversely, fungi from the Leotiomycetes, which host nearly one third of the known mycoviruses, only comprised 5% of the sequencing projects analyzed and 12% of the viruses identified.

Virus discovery favored the less data-rich classes of fungi. While Sordariomycetes and Eurotiomycetes host 73% of the viruses identified in this study, the discovery rate was 9% (37 viruses discovered in 415 datasets). Other than the sequencing samples from Lecanoromycetes

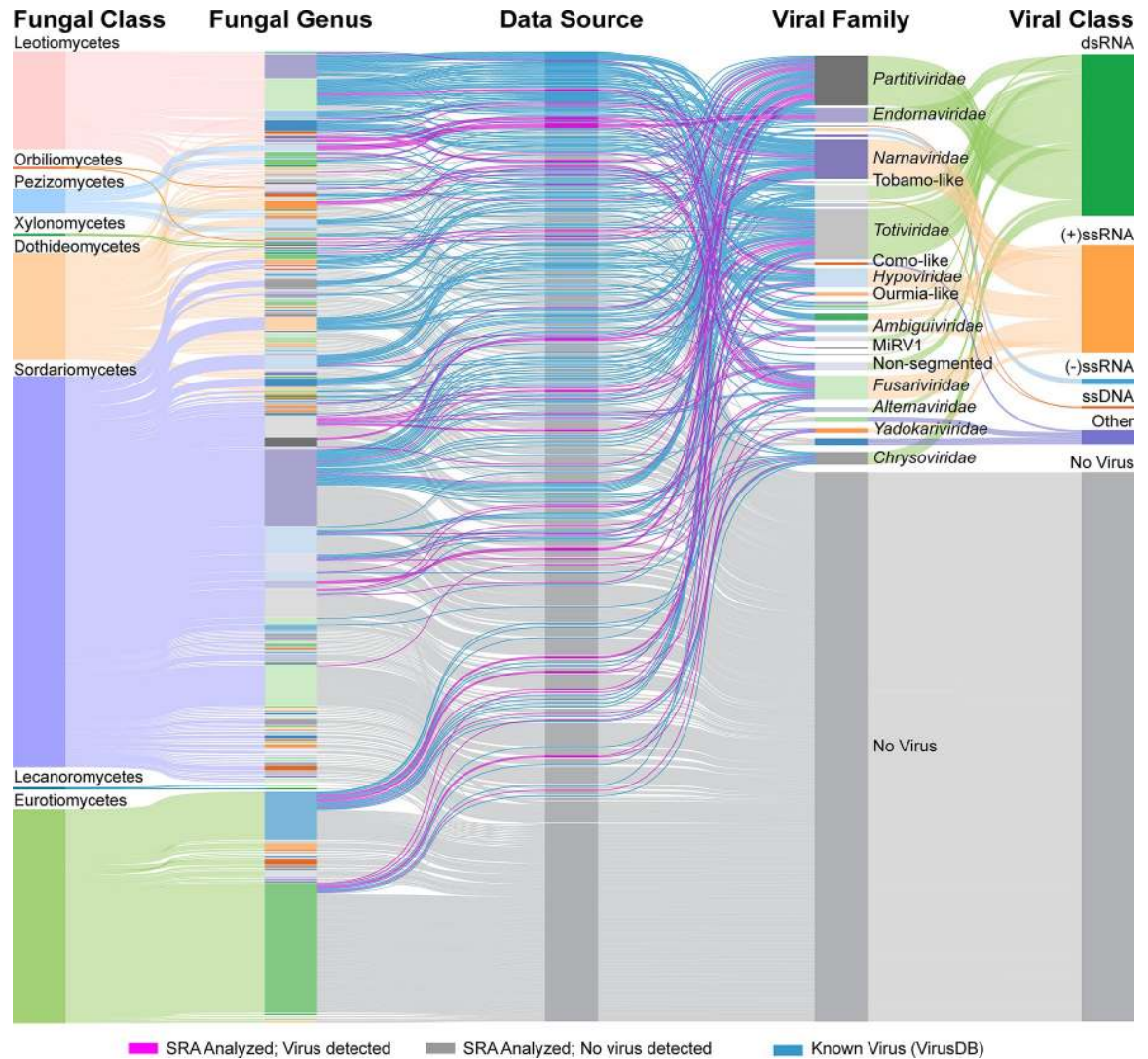


Fig 1. Sankey diagram summarizing the fungal host and virus state of the SRA samples queried in this study and recognized mycoviral sequences. The height of the bars in each column is indicative of the number of samples. Classification of fungal host is summarized in the two, leftmost columns. Column “Fungal Class”: fungal species are sorted into one of eight taxonomic classes, all genera from the same class are the same color; Column “Fungal Genus”: separates the fungi into their respective genera; colors used in this column are to highlight the individual genera but similar colors have no relation to one another. The third column, “Data Source” is color-coded by the origin of the sample; blue indicates a known mycovirus, and the colors pink and grey indicate an SRA sample was analyzed where either a virus sequence was detected (pink) or was not detected (grey). The two rightmost columns summarize the information related to virus taxonomy. Column 4: “Viral Family” organizes the samples that have either a known or discovered virus by viral family. Column 5: “Viral Class” summarizes the viral genomes by nucleic acid and/or strandedness. SRA sequencing samples without a virus detected are in the category “No Virus”.

<https://doi.org/10.1371/journal.pone.0219207.g001>

and Xylonomycetes where no viruses were detected from a total of four datasets, samples from the remaining classes yielded discovery rates ranging from 17% to 50%. Comparing these results to the distribution of known of mycoviruses revealed a 200% increase within the Pezizomycetes; previously three viral genomes had been described while six new mycoviral genomes were identified in this study. Similarly, while no mycoviruses from Orbiliomycetes were present within the known mycovirus dataset, one was identified in this analysis.

Classification of the viral sequences by the type of genome revealed that dsRNA viruses are the most abundant class within both datasets, with 58% and 59% belonging to those identified

in this study and the known set respectively. Further, characterization of the viruses by viral family demonstrated similar patterns across the new and known virus datasets, as members of the *Totiviridae* and *Partitiviridae* families are the most prevalent, followed by viruses belonging to the *Endornaviridae* and *Chrysoviridae* families. The remaining 42% of viruses identified in this study are classified as positive-strand, single-stranded RNA viruses ((+)ssRNA), where the most common family observed, with 10 genomes, was the proposed family *Fusariviridae*.

Virus discovery did not appear to be affected by selection steps used during library preparation; of the 38 libraries where this information was available, 36 samples sequenced poly(A)-purified RNAs. Among these 36 libraries, at least one (+)ssRNA virus was discovered in 16 samples, while 17 samples yielded at least one dsRNA virus. The remaining three libraries contained both (+)ssRNA and dsRNA viruses. Therefore, despite dsRNA viral genomes lacking a poly(A) tail, they were easily detectable within these datasets.

An in-depth analysis and characterization of each of the 59 viral sequences is described below (Table 1). Viral genomes have been divided into two major categories: dsRNA viruses and (+)ssRNA viruses then further separated by viral family and, where applicable, genus. In total, viruses were identified from eight of the currently recognized virus families and as such are described within the context of the expected characteristics. The remaining viruses form groups with other recently identified mycoviruses leading to the characterization of new genera within existing mycoviral families and as well as new mycoviral families. A common characteristic used to evaluate new viral genomic sequences for submission to the International Committee on Taxonomy of Viruses (ICTV) is the percent identity between the new sequences and existing sequences. As such this criteria is evaluated for the viruses that belong with existing families recognized by the ICTV. For viruses that fall into other groupings, a similar methodology is used thus providing context for relationships both within the newly described groups and between the new and existing groups.

Double-stranded RNA (dsRNA) viruses

A total of 34 viruses were categorized as dsRNA viruses due the identity of the top BLASTP match. To further characterize the relationship of these sequences, a phylogenetic tree based on the RdRP amino acid sequences from the 34 identified viruses and the top BLASTP matches was constructed (Fig 2). Globally, five groups are visible, where three are known viral families (*Totiviridae*, *Partitiviridae*, and *Chrysoviridae*) and two represent potentially new families. Properties characteristic of each group of viruses are described in detail below.

Non-segmented RNA virus. Recently, dsRNA viruses have been identified from various fungi that encode two ORFs within the single-segment genome, where the RdRP sequences have more similarity to the RdRPs of partitiviruses (multi-segment genomes) than totiviruses (mono-segment genomes). The virus *Beauveria bassiana non-segmented dsRNA virus 1* (BbNRV1), assembled using the RNA-seq data from a transcriptome experiment of the highly virulent isolate *B. bassiana* ARSEF 8028 [9], groups with these previously identified viruses (Fig 2). A second virus from this group was also identified in this analysis: *Colletotrichum higginsianum non-segmented RNA virus 1* (ChNRV1); this virus served as an internal control for the virus discovery pipeline as it was previously identified via a similar approach from two different *Colletotrichum higginsianum* RNA-seq datasets [8].

BbNRV1 and ChNRV1, together with four previously identified viruses, form a strongly supported group that is separate from other characterized fungal viruses (Fig 2). It has been noted by multiple researchers that these fungal viruses share similarities with a family of plant viruses, the *Amalgaviridae*, specifically the mono-segmented genome structure and partitiviruses-like RdRP sequence. Therefore, a phylogenetic analysis of the RdRP protein sequence

Table 1. Summary of mycoviral genomes identified from fungal RNA-seq datasets.

| Fungal Host | Virus Family | Virus Name | Virus Abbrev | Virus Class | BioProject and SRA ID (s) | Citation for Data ^a | Segment Number | Length (nt) | Nucleotide Accession | Protein ID | Length (aa) |
|--|------------------------|---|--------------|--------------|---------------------------|--------------------------------|----------------|-------------|----------------------|--------------|-------------|
| <i>Acidomyces richmondensis</i> BFW | Tobamo-like virus | <i>Acidomyces richmondensis tobamo-like virus 1</i> | ArTIV1 | (+) ssRNA | PRJNA250470 (SRR1797521) | JGI (SWS) | 1 | 10291 | MK279511 | ORF1-RdRP | 2182 |
| | | | | | | | | | | ORF3 | 859 |
| | | | | | | | | | | ORF4 | 334 |
| <i>Aspergillus ellipticus</i> strain CBS 707.79 | <i>Fusariviridae</i> | <i>Aspergillus ellipticus fusarivirus 1</i> | AeFV1 | (+) ssRNA | PRJNA250911 (SRR1799565) | JGI (SEB) | 1 | 6253 | MK279500 | RdRP | 1556 |
| | | | | | | | | | | ORF2 | 482 |
| <i>Aspergillus heteromorphus</i> isolate CBS 117.55 | <i>Alternaviridae</i> | <i>Aspergillus heteromorphus alternavirus 1</i> | AheAV1 | dsRNA | PRJNA250969 (SRR1799577) | JGI (SEB) | 1 | 3576 | MK279437 | RdRP | 1124 |
| | | | | | | | 2 | 2742 | MK279438 | Coat Protein | 833 |
| | | | | | | | 3 | 2427 | MK279439 | Hyp. Protein | 727 |
| <i>Aspergillus homomorphus</i> strain CBS 101889 | <i>Totiviridae</i> | <i>Aspergillus homomorphus totivirus 1</i> | AhoTV1 | dsRNA | PRJNA250984 (SRR1799578) | JGI (SEB) | 1 | 5147 | MK279489 | Coat Protein | 780 |
| | | | | | | | | | | RdRP | 826 |
| <i>Aspergillus homomorphus</i> strain CBS 101889 | <i>Yadokariviridae</i> | <i>Aspergillus homomorphus yadokarivirus 1</i> | AhoYV1 | (+) ssRNA | PRJNA250984 (SRR1799578) | JGI (SEB) | 1 | 3621 | MK279487 | RdRP | 963 |
| <i>Aspergillus neoniger</i> isolate CBS 115656 | Ourmia-like virus | <i>Aspergillus neoniger ourmia-like virus 1</i> | AnOIV1 | (+) ssRNA | PRJNA250996 (SRR1801411) | JGI (SEB) | 1 | 2708 | MK279481 | RdRP | 769 |
| <i>Beauveria bassiana</i> strain ARSEF 8028 | Unclassified dsRNA | <i>Beauveria bassiana non-segmented virus 1</i> | BbNRV1 | dsRNA | PRJNA260878 (SRR3269778) | [9] | 1 | 3170 | MK279499 | ORF1 | 315 |
| | | | | | | | | | | RdRP | 590 |
| <i>Beauveria bassiana</i> strain GXsk1011 | <i>Chrysoviridae</i> | <i>Beauveria bassiana chrysovirus 1</i> | BbCV1 | dsRNA | PRJNA306902 (SRR3043102) | [10] | 1 | 3478 | MK279433 | RdRP | 1115 |
| | | | | | | | 2 | 3143 | MK279434 | Coat Protein | 976 |
| | | | | | | | 3 | 3069 | MK279435 | Hyp. Protein | 964 |
| | | | | | | | 4 | 2770 | MK279436 | Hyp. Protein | 851 |
| Fungal Host | Virus Family | Virus Name | Virus Abbrev | Virus Class | BioProject and SRA ID (s) | Citation for Data ^a | Segment Number | Length (nt) | Nucleotide Accession | Protein ID | Length (aa) |
| <i>Clohesyomyces aquaticus</i> strain CBS 115471 | <i>Partitiviridae</i> | <i>Clohesyomyces aquaticus partitivirus 1</i> | CaPV1 | dsRNA | PRJNA372822 (SRR5494002) | [11] | 1 | 1873 | MK279448 | RdRP | 578 |
| | | | | | | | 2 | 1850 | MK279449 | Coat Protein | 536 |
| <i>Colletotrichum caudatum</i> strain CBS 131602 | <i>Totiviridae</i> | <i>Colletotrichum caudatum totivirus 1</i> | CcTV1 | dsRNA | PRJNA262373 (SRR5145569) | JGI (JC) | 1 | 5125 | MK279490 | Coat Protein | 787 |
| | | | | | | | | | | RdRP | 825 |
| <i>Colletotrichum eremochloae</i> isolate CBS 129661 | <i>Partitiviridae</i> | <i>Colletotrichum eremochloae partitivirus 1</i> | CePV1 | dsRNA | PRJNA262412 (SRR5166050) | JGI (JC) | 1 | 1815 | MK279452 | RdRP | 573 |
| | | | | | | | 2 | 1696 | MK279453 | Coat Protein | 494 |
| <i>Colletotrichum eremochloae</i> isolate CBS129661 | <i>Totiviridae</i> | <i>Colletotrichum eremochloae totivirus 1</i> | CeTV1 | dsRNA | PRJNA262412 (SRR5166050) | JGI (JC) | 1 | 5364 | MK279491 | Coat Protein | 757 |
| | | | | | | | | | | RdRP | 862 |
| <i>Colletotrichum falcatum</i> strain CoC 671 | <i>Mitovirus</i> | <i>Colletotrichum falcatum mitovirus 1</i> | CfMV1 | (+) ssRNA | PRJNA272832 (SRR1765657) | JGI (JC) | 1 | 2283 | MK279482 | RdRP | 703 |
| <i>Colletotrichum navitas</i> strain VF38c | <i>Totiviridae</i> | <i>Colletotrichum navitas totivirus 1</i> | CnTV1 | dsRNA | PRJNA262225 (SRR5166338) | JGI (JC) | 1 | 5070 | MK279492 | Coat Protein | 757 |
| | | | | | | | | | | RdRP | 831 |

(Continued)

Table 1. (Continued)

| | | | | | | | | | | | |
|--|----------------|---|--------------|--------------|---------------------------|--------------------------------|----------------|-------------|----------------------|--------------|-------------|
| <i>Colletotrichum zoysiae</i> strain MAFF235873 | Totiviridae | Colletotrichum zoysiae totivirus 1 | CzTV1 | dsRNA | PRJNA262216 (SRR5166062) | JGI (JC) | 1 | 5174 | MK279493 | Coat Protein | 783 |
| | | | | | | | | | | RdRP | 831 |
| <i>Cryphonectria parasitica</i> strain cpku80 | Partitiviridae | Penicillium aurantiogriseum partitivirus 1-cp | PaPV1-cp | dsRNA | PRJNA369604 (SRR5235483) | [12] | 1 | 1800 | MK279462 | RdRP | 573 |
| | | | | | | | 2 | 1670 | MK279463 | Coat Protein | 494 |
| <i>Delitschia confertasporea</i> strain ATCC 74209 | Partitiviridae | Delitschia confertasporea partitivirus 1 | DcPV1 | dsRNA | PRJNA250740 (SRR3440303) | JGI (JS) | 1 | 1899 | MK279446 | RdRP | 571 |
| | | | | | | | 2 | 1645 | MK279447 | Coat Protein | 501 |
| Fungal Host | Virus Family | Virus Name | Virus Abbrev | Virus Class | BioProject and SRA ID (s) | Citation for Data ^a | Segment Number | Length (nt) | Nucleotide Accession | Protein ID | Length (aa) |
| <i>Drechslerella stenobrocha</i> strain 248 | Partitiviridae | Drechslerella stenobrocha partitivirus 1 | DsPV1 | dsRNA | PRJNA236481 (SRR1145648) | [13] | 1 | 1935 | MK279440 | RdRP | 586 |
| | | | | | | | 2 | 1721 | MK279441 | Coat Protein | 521 |
| <i>Fusarium graminearum</i> strain HN10 | Hypoviridae | Fusarium graminearum hypovirus 1-HN10 | FgHV1-HN10 | (+) ssRNA | PRJNA263651 (SRR4445678) | [14] | 1 | 13035 | MK279472 | ORFA | 176 |
| | | | | | | | | | | ORFB | 3705 |
| <i>Fusarium poae</i> isolate 2516 | Partitiviridae | Fusarium poae partitivirus 1-2516 | FpPV1-2516 | dsRNA | PRJNA319914 (SRR3953125) | [15] | 1 | 2271 | MK279442 | RdRP | 701 |
| | | | | | | | 2 | 2199 | MK279443 | Coat Protein | 637 |
| <i>Gaeumannomyces tritici</i> strain GGT-007 | Fusariviridae | Gaeumannomyces tritici fusarivirus 1 | GtFV1 | (+) ssRNA | PRJNA268052 (SRR1664730) | [16] | 1 | 6332 | MK279501 | RdRP | 1525 |
| | | | | | | | | | | ORF2 | 504 |
| <i>Gaeumannomyces tritici</i> strain GGT-007 | Partitiviridae | Gaeumannomyces tritici partitivirus 1 | GtPV1 | dsRNA | PRJNA268052 (SRR1664730) | [16] | 1 | 1881 | MK279464 | RdRP | 578 |
| | | | | | | | 2 | 1739 | MK279465 | Coat Protein | 494 |
| <i>Gaeumannomyces tritici</i> strain GGT-007 | Partitiviridae | Gaeumannomyces tritici partitivirus 2 | GtPV2 | dsRNA | PRJNA268052 (SRR1664730) | [16] | 1 | 2058 | MK279466 | RdRP | 617 |
| | | | | | | | 2 | 1798 | MK279467 | Coat Protein | 487 |
| <i>Grosmannia clavigera</i> strain KW1407 | Partitiviridae | Grosmannia clavigera partitivirus 1 | GcPV1 | dsRNA | PRJNA184372 (SRR636711) | [17] | 1 | 2040 | MK279468 | RdRP | 599 |
| | | | | | | | 2 | 1703 | MK279469 | Coat Protein | 479 |
| <i>Gyromitra esculenta</i> strain CBS 101906 | Endornaviridae | Gyromitra esculenta endornavirus 1 | GeEV1 | dsRNA | PRJNA372840 (SRR5491178) | JGI (JS) | 1 | 14584 | MK279476 | Polyprotein | 4846 |
| <i>Hortaea werneckii</i> strain EXF-2000 | Totiviridae | Hortaea werneckii totivirus 1 | HwTV1 | dsRNA | PRJNA356640 (SRR5086622) | [18] | 1 | 4580 | MK279498 | Coat Protein | 681 |
| | | | | | | | | | | RdRP | 805 |
| Fungal Host | Virus Family | Virus Name | Virus Abbrev | Virus Class | BioProject and SRA ID (s) | Citation for Data ^a | Segment Number | Length (nt) | Nucleotide Accession | Protein ID | Length (aa) |
| <i>Loramyces juncicola</i> strain ATCC 46458 | Mitovirus | Loramyces juncicola mitovirus 1 | LjMV1 | (+) ssRNA | PRJNA372853 (SRR5487643) | JGI (JS) | 1 | 2416 | MK279483 | RdRP | 688 |
| <i>Magnaporthe grisea</i> strain M82 | Partitiviridae | Magnaporthe grisea partitivirus 1 | MgPV1 | dsRNA | PRJNA269089 (SRR1695911) | [19] | 1 | 1725 | MK279458 | RdRP | 539 |
| | | | | | | | 2 | 1530 | MK279459 | Coat Protein | 437 |
| <i>Morchella importuna</i> strain SCYDJ1-A1 | Endornaviridae | Morchella importuna endornavirus 1 | MiEV1 | dsRNA | PRJNA372858 (SRR5487664) | JGI (JS) | 1 | 16495 | MK279477 | Polyprotein | 5447 |
| <i>Morchella importuna</i> strain SCYDJ1-A1 | Endornaviridae | Morchella importuna endornavirus 2 | MiEV2 | dsRNA | PRJNA372858 (SRR5487664) | JGI (JS) | 1 | 15394 | MK279478 | Polyprotein | 4956 |

(Continued)

Table 1. (Continued)

| | | | | | | | | | | | |
|---|------------------------|---|--------------|--------------|---------------------------|--------------------------------|----------------|-------------|----------------------|--------------|-------------|
| <i>Morchella importuna</i> strain SCYDJ1-A1 | <i>Endornaviridae</i> | <i>Morchella importuna</i> endornavirus 3 | MiEV3 | dsRNA | PRJNA372858 (SRR5487664) | JGI (JS) | 1 | 14296 | MK279479 | Polyprotein | 4734 |
| <i>Morchella importuna</i> strain SCYDJ1-A1 | <i>Fusariviridae</i> | <i>Morchella importuna</i> fusarivirus 1 | MiFV1 | (+) ssRNA | PRJNA372858 (SRR5487664) | JGI (JS) | 1 | 7835 | MK279502 | RdRP | 1480 |
| | | | | | | | | | | ORF2 | 381 |
| | | | | | | | | | | ORF3 | 622 |
| <i>Morchella importuna</i> strain SCYDJ1-A1 | Unclassified ssRNA | <i>Morchella importuna</i> RNA virus 1 | MiRV1 | ssRNA | PRJNA372858 (SRR5487664) | JGI (JS) | 1 | 10099 | MK279480 | Polyprotein | 3295 |
| <i>Neurospora discreta</i> strain FGSC8579 | <i>Fusariviridae</i> | <i>Neurospora discreta</i> fusarivirus 1 | NdFV1 | (+) ssRNA | PRJNA257829 (SRR1539773) | [20] | 1 | 6648 | MK279503 | RdRP | 1526 |
| | | | | | | | | | | ORF2 | 522 |
| <i>Ophiocordyceps sinensis</i> strain 1229 | <i>Mitovirus</i> | <i>Ophiocordyceps sinensis</i> mitovirus 1 | OsMV1 | (+) ssRNA | PRJNA292632 (SRR2533613) | [21] | 1 | 2386 | MK279484 | RdRP | 698 |
| Fungal Host | Virus Family | Virus Name | Virus Abbrev | Virus Class | BioProject and SRA ID (s) | Citation for Data ^a | Segment Number | Length (nt) | Nucleotide Accession | Protein ID | Length (aa) |
| <i>Ophiocordyceps sinensis</i> strain 1229 | <i>Mitovirus</i> | <i>Ophiocordyceps sinensis</i> mitovirus 2 | OsMV2 | (+) ssRNA | PRJNA292632 (SRR2533613) | [21] | 1 | 2439 | MK279485 | RdRP | 716 |
| <i>Penicillium brasilianum</i> strain MG11 | <i>Partitiviridae</i> | <i>Penicillium brasilianum</i> partitivirus 1 | PbPV1 | dsRNA | PRJEB7514 (ERR677271) | [22] | 1 | 1791 | MK279470 | RdRP | 569 |
| | | | | | | | 2 | 1518 | MK279471 | Coat Protein | 453 |
| <i>Penicillium digitatum</i> KH8 | <i>Totiviridae</i> | <i>Penicillium digitatum</i> totivirus 2 | PdTV2 | dsRNA | PRJNA254400 (SRR1557148) | [23] | 1 | 5184 | MK279494 | Coat Protein | 807 |
| | | | | | | | | | | RdRP | 838 |
| <i>Penicillium digitatum</i> KH8 | <i>Totiviridae</i> | <i>Penicillium digitatum</i> totivirus 1-KH8 | PdTV1-KH8 | dsRNA | PRJNA254400 (SRR1557148) | [23] | 1 | 5168 | MK279495 | Coat Protein | 776 |
| | | | | | | | | | | RdRP | 825 |
| <i>Penicillium digitatum</i> strain HS-F6 | <i>Partitiviridae</i> | <i>Penicillium digitatum</i> partitivirus 1 | PdPV1 | dsRNA | PRJNA352307 (SRR4851221) | CCNU (QC) | 1 | 1781 | MK279456 | RdRP | 539 |
| | | | | | | | 2 | 1526 | MK279457 | Coat Protein | 434 |
| <i>Penicillium digitatum</i> strain HS-F6 | <i>Yadokariviridae</i> | <i>Penicillium digitatum</i> yadokarivirus 1 | PdYV1 | (+) ssRNA | PRJNA352307 (SRR4851221) | CCNU (QC) | 1 | 3507 | MK279488 | RdRP | 987 |
| <i>Penicillium raistrickii</i> strain ATCC 10490 | <i>Chrysoviridae</i> | <i>Penicillium raistrickii</i> chrysovirus 1 | PrCV1 | dsRNA | PRJNA250734 (SRR1801283) | JGI (DS) | 1 | 3501 | MK279429 | RdRP | 1116 |
| | | | | | | | 2 | 3120 | MK279430 | Coat Protein | 983 |
| | | | | | | | 3 | 2955 | MK279431 | Hyp. Protein | 901 |
| | | | | | | | 4 | 2980 | MK279432 | Hyp. Protein | 847 |
| <i>Periconia macrospinoso</i> strain DSE2036 | <i>Ambiguiviridae</i> | <i>Periconia macrospinoso</i> ambiguivirus 1 | PmAV1 | ssRNA | PRJNA262386 (SRR5153210) | [24] | 1 | 3566 | MK279507 | ORF1 (HP) | 336 |
| | | | | | | | | | | RdRP | 459 |
| <i>Phyllosticta citriasiana</i> strain CBS 120486 | <i>Partitiviridae</i> | <i>Phyllosticta citriasiana</i> partitivirus 1 | PcPV1 | dsRNA | PRJNA250394 (SRR3314103) | JGI (JS) | 1 | 1700 | MK279450 | RdRP | 535 |
| | | | | | | | 2 | 1504 | MK279451 | Coat Protein | 421 |
| Fungal Host | Virus Family | Virus Name | Virus Abbrev | Virus Class | BioProject and SRA ID (s) | Citation for Data ^a | Segment Number | Length (nt) | Nucleotide Accession | Protein ID | Length (aa) |
| <i>Pseudogymnoascus destructans</i> strain 20631-21 | <i>Partitiviridae</i> | <i>Pseudogymnoascus destructans</i> partitivirus-pa | PdPVpa | dsRNA | PRJNA66121 (SRR254216) | Broad Institute | 1 | 1754 | MK279444 | RdRP | 539 |
| | | | | | | | 2 | 1562 | MK279445 | Coat Protein | 434 |

(Continued)

Table 1. (Continued)

| | | | | | | | | | | | |
|--|-----------------------|--|--------------|-------------|--|---------------------------------|----------------|-------------|----------------------|--------------|-------------|
| <i>Rutstroemia firma</i> strain CBS 115.86 | <i>Fusariviridae</i> | <i>Rutstroemia firma</i> fusarivirus 1 | RfFV1 | (+) ssRNA | PRJNA372878 (SRR5689333) | JGI (JS) | 1 | 6641 | MK279504 | RdRP | 1558 |
| | | | | | | | | | | ORF2 | 545 |
| <i>Sclerotinia homoeocarpa</i> isolate LT11 | <i>Fusariviridae</i> | <i>Sclerotinia homoeocarpa</i> fusarivirus 1 | ShFV1 | (+) ssRNA | PRJNA167556 (SRR515157) | [25] | 1 | 7171 | MK279505 | RdRP | 1718 |
| | | | | | | | | | | ORF2 | 589 |
| <i>Sclerotinia homoeocarpa</i> isolate LT11 | <i>Hypoviridae</i> | <i>Sclerotinia homoeocarpa</i> hypovirus 1 | ShHV1 | (+) ssRNA | PRJNA167556 (SRR515157) | [25] | 1 | 12370 | MK279473 | Polyprotein | 3604 |
| <i>Setosphaeria turcica</i> strain 28A | <i>Ambiguiviridae</i> | <i>Setosphaeria turcica</i> ambiguiavirus 1 | StAV1 | ssRNA | PRJNA250530 (SRR1587420) | CU (GT) | 1 | 3478 | MK279508 | ORF1 (HP) | 344 |
| | | | | | | | | | | RdRP | 496 |
| <i>Setosphaeria turcica</i> strain 28A | <i>Hypoviridae</i> | <i>Setosphaeria turcica</i> hypovirus 1 | StHV1 | (+) ssRNA | PRJNA250530 (SRR1587420) | CU (GT) | 1 | 9069 | MK279474 | Polyprotein | 2751 |
| <i>Setosphaeria turcica</i> strain 28A | <i>Mitovirus</i> | <i>Setosphaeria turcica</i> mitovirus 1 | StMV1 | (+) ssRNA | PRJNA250530 (SRR1587420) | CU (GT) | 1 | 2595 | MK279486 | RdRP | 652 |
| <i>Thelebolus microsporus</i> ATCC 90970 | <i>Partitiviridae</i> | <i>Thelebolus microsporus</i> partitivirus 1 | TmPV1 | dsRNA | PRJNA372886 (SRR5487623) | PNNL (JM) | 1 | 1696 | MK279460 | RdRP | 533 |
| | | | | | | | 2 | 1531 | MK279461 | Coat Protein | 444 |
| <i>Thelebolus microsporus</i> strain ATCC 90970 | <i>Totiviridae</i> | <i>Thelebolus microsporus</i> totivirus 1 | TmTV1 | dsRNA | PRJNA372886 (SRR5487623) | PNNL (JM) | 1 | 5146 | MK279496 | Coat Protein | 763 |
| | | | | | | | | | | RdRP | 829 |
| <i>Tolypocladium ophioglossoides</i> strain CBS 100239 | <i>Totiviridae</i> | <i>Tolypocladium ophioglossoides</i> totivirus 1 | ToTV1 | dsRNA | PRJNA292830 (SRR2179765) | [26] | 1 | 5264 | MK279497 | Coat Protein | 787 |
| | | | | | | | | | | RdRP | 831 |
| Fungal Host | Virus Family | Virus Name | Virus Abbrev | Virus Class | BioProject and SRA ID (s) | Citation for Data ^a | Segment Number | Length (nt) | Nucleotide Accession | Protein ID | Length (aa) |
| <i>Trichoderma asperellum</i> strain CBS 131938 | <i>Hypoviridae</i> | <i>Trichoderma asperellum</i> hypovirus 1 | TaHV1 | (+) ssRNA | PRJNA261111 (SRR1575447) | [27] | 1 | 14211 | MK279475 | Polyprotein | 4176 |
| <i>Trichoderma citrinoviride</i> strain FP-102208 | <i>Partitiviridae</i> | <i>Trichoderma citrinoviride</i> partitivirus 1 | TcPV1 | dsRNA | PRJNA304029 (SRR2961293) | [28] | 1 | 2284 | MK279454 | RdRP | 727 |
| | | | | | | | 2 | 2297 | MK279455 | Coat Protein | 663 |
| <i>Trichoderma harzianum</i> strain TR274 | <i>Ambiguiviridae</i> | <i>Trichoderma harzianum</i> ambiguiavirus 1 | ThAV1 | ssRNA | PRJNA216008 (SRR976276-8) | [29] | 1 | 3911 | MK279509 | ORF1 (HP) | 464 |
| | | | | | | | | | | RdRP | 493 |
| <i>Verticillium longisporum</i> isolate 43 | <i>Ambiguiviridae</i> | <i>Verticillium longisporum</i> ambiguiavirus 1 | VlAV1 | ssRNA | PRJNA308558 (SRR3102592) | EMAUG (KH) | 1 | 2798 | MK279510 | ORF1 (HP) | 256 |
| | | | | | | | | | | RdRP | 494 |
| <i>Zyloseptoria tritici</i> strain IPO323 | <i>Fusariviridae</i> | <i>Zyloseptoria tritici</i> fusarivirus 1 | ZtFV1 | (+) ssRNA | PRJNA179083 (SRR612175) PRJEB8798 (ERR789231) PRJNA237967 (SRR1167717) PRJNA253135 (SRR1427070) | UE (DS) [30] [31] [32] | 1 | 5969 | MK279506 | RdRP | 1491 |
| | | | | | | | | | | ORF2 | 499 |

^a Abbreviations corresponding to the contact person for unpublished datasets

JGI (SWS): Joint Genome Institute, Steven W. Singer

JGI (SEB): Joint Genome Institute, Scott E. Baker

JGI (JC): Joint Genome Institute, Jo Ann Crouch

JGI (JS): Joint Genome Institute, Joseph Spatafora

CCNU (QC): Central China Normal University, Qianwen Cao

JGI (DG): Joint Genome Institute, Dave Greenshields

CU (GT): Cornell University, Gillian Turgeon

PNNL (JM): Pacific Northwest National Laboratory, Jon Magnuson

EMAUG (KH): Ernst-Moritz-Arndt University of Greifswald, Katharina Hoff

UE (DS): University of Exeter, David Studholme

<https://doi.org/10.1371/journal.pone.0219207.t001>

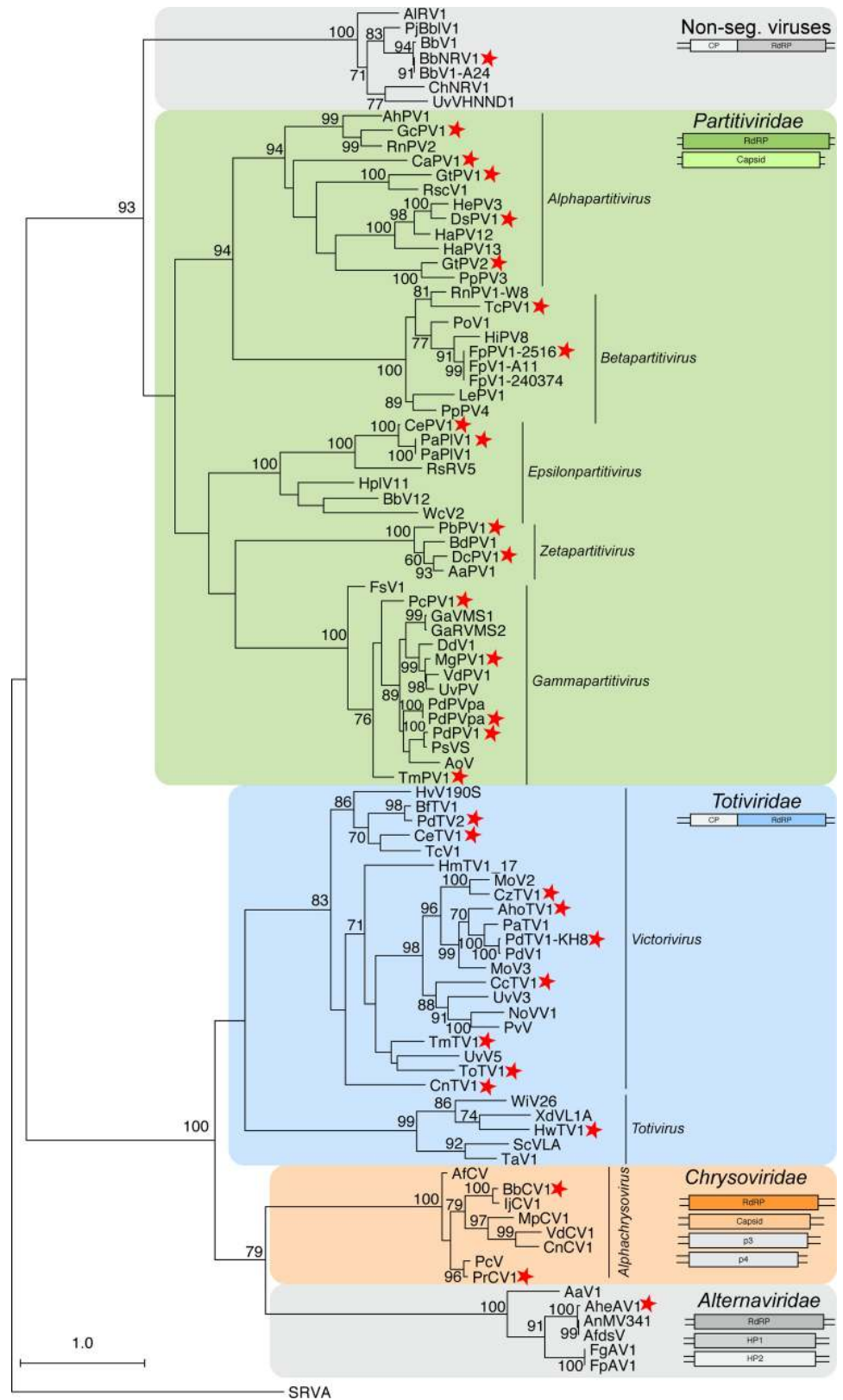


Fig 2. Phylogram of the RdRP-containing ORF for viruses classified as dsRNA viruses. A maximum likelihood phylogenetic tree was generated with RaxML, using the LG+G+I amino acid substitution model. Scale bar represents

1.0 amino acid substitutions per site and numbers at the nodes indicate bootstrap support over 70% (1000 replicates). The sequence from Simian Rotavirus A (SRVA) serves as the outgroup. Three officially recognized families are present: *Partitiviridae*, *Chrysoviridae*, and *Totiviridae* and are highlighted with a green, orange, or blue box respectively. Gray boxes surround families currently unrecognized. The structure of the genomic segment(s) for each highlighted group is depicted below the family name. Genera within the *Partitiviridae*, *Chrysoviridae* and *Totiviridae* family are indicated. Viruses identified in this study are noted with a red star. RdRP: RNA-dependent RNA Polymerase; CP: Coat Protein; HP: Hypothetical Protein.

<https://doi.org/10.1371/journal.pone.0219207.g002>

was performed which demonstrates that, while the mono-segmented genome structure is shared, this new clade of fungal viruses is also distinct from the plant amalgaviruses (Fig 3A). Indeed, the amalgavirus group of plant viruses branches off before the BbNRV1-related and *Partitivirus* clades. A percent identity matrix generated with RdRP and CP coding sequences further demonstrates a separation between this new group and existing virus families (Fig 3B). For the RdRP pairwise comparisons, 50% identity easily differentiates the three clades from

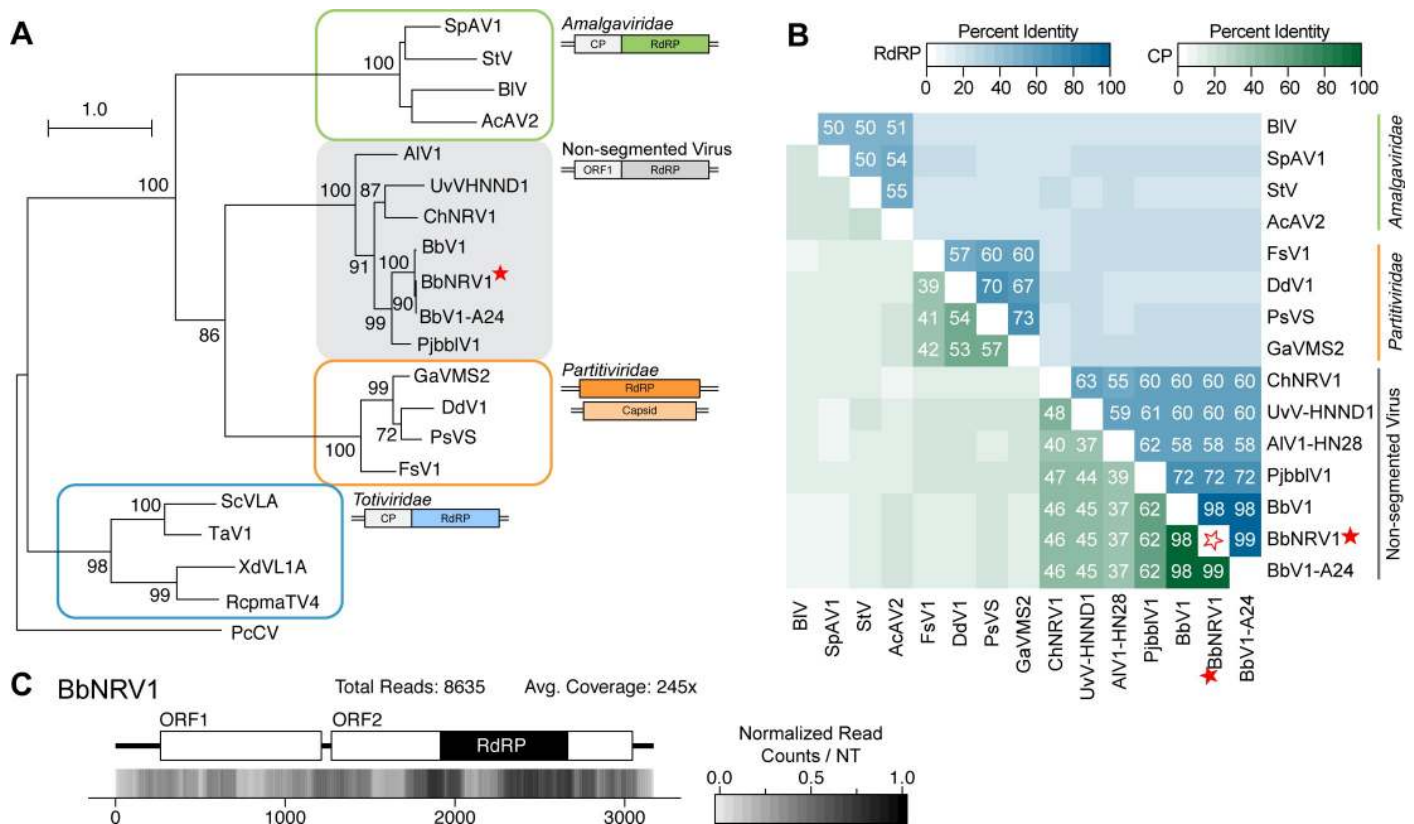


Fig 3. Analyses of identified novel non-segmented RNA viral genomes in relation to Amalgaviridae and Partitiviridae. (A) Phylogram of the ORF containing the RdRP motif for members of the new mycovirus group along with select partitiviruses, totiviruses and amalgaviruses; sequence from *Penicillium chrysogenum virus* (PcCV) serves as the outgroup. Full names and accession numbers for protein sequences used in the alignment are in S2 Table. Scale bar on the phylogenetic tree represents 1.0 amino acid substitutions per site and numbers at the nodes indicate bootstrap support over 70% (1000 replicates). Diagrams illustrating the genome structure and predicted ORFs are present for each group of viruses. (B) Percent identity matrix, generated by Clustal-Omega 2.1, of new mycoviral family, partitiviruses, and amalgaviruses. The top half of the matrix, in the blue scale, is the percent identity of the RdRP protein coding sequence and the bottom half of the matrix, in the green scale, is the percent identity of the putative CP. For sake of clarity, the 100% identity along the diagonal has been removed and outlined stars are used to note junction point for the row and column of the viruses identified. Only RdRP values $\geq 50\%$ and CP values $\geq 30\%$ are indicated. (C) Density plot of read counts per nucleotide across the viral genome for BbNRV1. Graphical depiction of the genome structure, ORFs and predicted motifs are on along the top and a heatmap of normalized read counts is on the bottom. Read counts are normalized to a 0 – 1 scale, where the maximum read count for a given genome equals 1. Total reads mapped to the virus genome and average read coverage are noted. The red star indicates the virus identified in this study. RdRP: RNA-dependent RNA Polymerase; CP: Coat Protein; ORF: Open Reading Frame.

<https://doi.org/10.1371/journal.pone.0219207.g003>

each other. Additionally, the CP sequence comparisons demonstrate shared similarities within the three clades and less than 20% similarity is shared between the clades.

The genome structure of viruses within this group consists of two predicted ORFs surrounded by variable length UTRs (Fig 3C). The 5'UTR lengths fall into two distinct ranges: less than 40 nt for *Penicillium janczewskii Beauveria bassiana-like virus 1* (PjBbIV1) [4], ChNRV1 [8], and UvHNND1 [33], or between 260 to 320 nt for the three viruses isolated from *B. bassiana* [34,35] and *Alternaria longipes dsRNA virus 1* (ALV1-HN28) [36]. The protein encoded by ORF1 is predicted to be from 314 to 394 aa and in a +1 frame relative to the protein encoded by ORF2. The intergenic sequence lengths are generally less than 100 nt, with the exception of ALV1-HN28. ORF2 nucleotide sequences range from 1578 to 1773 nt, and encoded proteins, which encode an RdRP domain, range in size from 525 to 590 aa in length. A slippery site sequence identified between ORF1 and ORF2 of ChNRV1 suggested that an ORF1-ORF2 fusion protein may be made via a -1 ribosomal frameshift; indeed, trypsin digest followed by mass spectrometry of a protein near the predicted weight of the fusion protein returned signatures for both the ORF1 and ORF2 protein sequences [8]. Numerous attempts to isolate virus particles definitely produced by a virus within this group remain unsuccessful [4,8,34].

Due to the shared genome structure and partitivirus-like RdRP protein sequence, early viruses from this group were tentatively grouped with the plant *Amalgaviridae* family. However, with the identification of additional genome sequences, a more in depth phylogenetic analysis demonstrates the mycoviruses described here are separate. A new genus, “Unirnavirus”, has recently been suggested to reflect the single, unified genome segment that contains both ORFs [35]. As percent identity is a common criteria for species demarcation, the pairwise comparisons here provide possible thresholds for determining new species within this group of viruses; excluding the three viruses isolated from different strains of *B. bassiana*, a threshold of 75% identity in the RdRP protein sequence and 70% in the ORF1 coding sequence would be appropriate criteria to distinguish between the viruses isolated from unique fungal hosts.

Alternaviridae. Initially, a single segment encoding a single gene with an RdRP domain was identified from an *Aspergillus heteromorphus* sequencing sample (PRJNA250969; unpublished). BLAST alignment to ‘nr’ identified five related RdRP mycoviral sequences from viruses with multi-segmented genomes. *Fusarium graminearum alternavirus 1* (FgAV1) includes at least one additional genomic segment while *Alternaria alternata virus 1* (AaV1) [37] and *Aspergillus foetidus dsRNA mycovirus* (Afdsv) [38] each contain a total of four genomic segments. Using the additional sequences from these related viruses, two more contigs were identified from the *Aspergillus heteromorphus* Trinity assembly. Thus, the AheAV1 genome contains at least three genomic segments, the largest of which encodes the RdRP enzyme and the two smaller segments that encode proteins of unknown function (Table 1). Recently it has been proposed that a new family be established, “*Alternaviridae*” [38] we support the formation of this family, and the requisite genus (*Alternavirus*) and propose that AheAV1 is also a member of this new mycoviral family.

Previously it was observed that members of this clade group together, but separate from other families [37,38]. A phylogenetic analysis using the RdRP coding sequence demonstrates that the group containing AheAV1 forms a well-supported clade that appears related to the *Chrysoviridae* and *Totiviridae* families (Fig 2). Members of the *Totiviridae* family have single-segment genomes containing two ORFs, while members of the *Chrysoviridae* have multi-segment genomes. While the RdRP segment sizes are similar between alternaviruses and chrysoviruses (~3,500 bp), the remaining two to three segments are hundreds of base pairs shorter than those of chrysoviruses. Further, the proteins of unknown function have no predicted domains or shared sequence homology outside of this group of viruses. The three viral RdRP

sequences isolated from different species of *Aspergillus*, including AheAV1, are most closely related; similarly, the RdRP sequences from the two viruses from *Fusarium* species are most related (Fig 2). Indeed, a pairwise percent identity comparison of both the RdRP and HP1 protein sequences demonstrates the very high identity among the viruses from the two *Fusarium* and also from the two *Aspergillus* viruses (Fig 4A). This analysis also illustrates a clear delineation between alternaviruses and chrysovirus.

A common feature to viral genomes with multiple segments is the presence of a conserved sequence within the 5' UTRs of each fragment. Sequence conservation was described within the UTRs of the other putative alternaviruses, including AfV [38] and AaV1 [37]. Alignment of the 5'UTR sequences of the three genomic segments of AheAV1 demonstrates sequence conservation (Fig 4C). Alignment of the source RNA-seq reads to the viral genome segments reveals a high amount of total reads, with over four million reads mapped across the three segments (Fig 4B). This represents nearly 25% of the reads used for the analysis (reads that did not align to the fungal genome) and over 5% of the total reads from the sample. In an effort to identify a possible fourth genomic fragment for AheAV1, both the 5'UTR conserved sequence and the high read count were used as criteria to further query the Trinity assembled contigs however no additional contigs were identified.

An alignment of the RdRP protein sequences from the five alternavirus genomes reveals the presence of the expected eight conserved domains found within viral RdRP proteins (Fig 4D). Certain differences are shared among these viruses that are unique when compared to known viral sequences. Specifically, the triad within domain VI has an alanine (ADD) instead of the nearly universally conserved glycine (GDD).

Taken together, a number of criteria identify a new *Alternavirus*: multiple genome segments, where the largest encodes an RdRP protein, and the presence of an alanine within motif VI of the RdRP protein sequence. Classification of a new species within the genus *Alternavirus* remains to be established. To date, the distinguishing factors between viruses within this genus are the fungal host species and the number of genomic segments. While the fungal host species and number of genomic segments differ between AheAV1 and AfdV, the percent identity between the amino acid sequences from RNA1 and RNA2 is high, at 93% and 91% respectively. At this junction, it is not possible to determine if AheAV1 is a new species within the genus *Alternavirus*.

Totiviridae. Ten of the viral sequences identified in this study are similar to other known members of the *Totiviridae* family, nine within the genus *Victorivirus* and one within the genus *Totivirus*.

Totivirus. *Hortaea werneckii totivirus 1* (HwTV1), the first viral sequence described for this fungal species to our knowledge (PRJNA356640; [18]), is a member of the *Totivirus* genus (Fig 2). Species demarcation within *Totivirus* requires biological characterization of the virus within a host cell; however, isolation of a virus from a distinct host species and less than 50% sequence identity of the RdRP at the protein level is a proxy for identifying a separate species [39]. HwTV1 was identified from a new fungal host, *H. werneckii*, and shares less than 44% sequence similarity to the most closely related known totiviruses (S1A Fig) thus satisfying the proxy criteria for a new virus species. The RdRP protein sequence of HwTV1 shares the greatest similarity to two totiviruses identified within the fungus *Xanthophyllum dendrorhous*, XdV-L1-A and XdV-L1-B [40], along with the invertebrate virus *Wuhan insect virus 26* (WiV26), isolated from RNA-sequencing data of a mixed sample of flea and ants isolated in Hubei, China [41].

A common feature to members of the *Totivirus* genus is the presence of two partially overlapping open reading frames that produce a fusion protein via a -1 frameshift event with the 5' ORF encoding the major capsid protein and the 3' ORF encoding an RdRP. Characteristics of

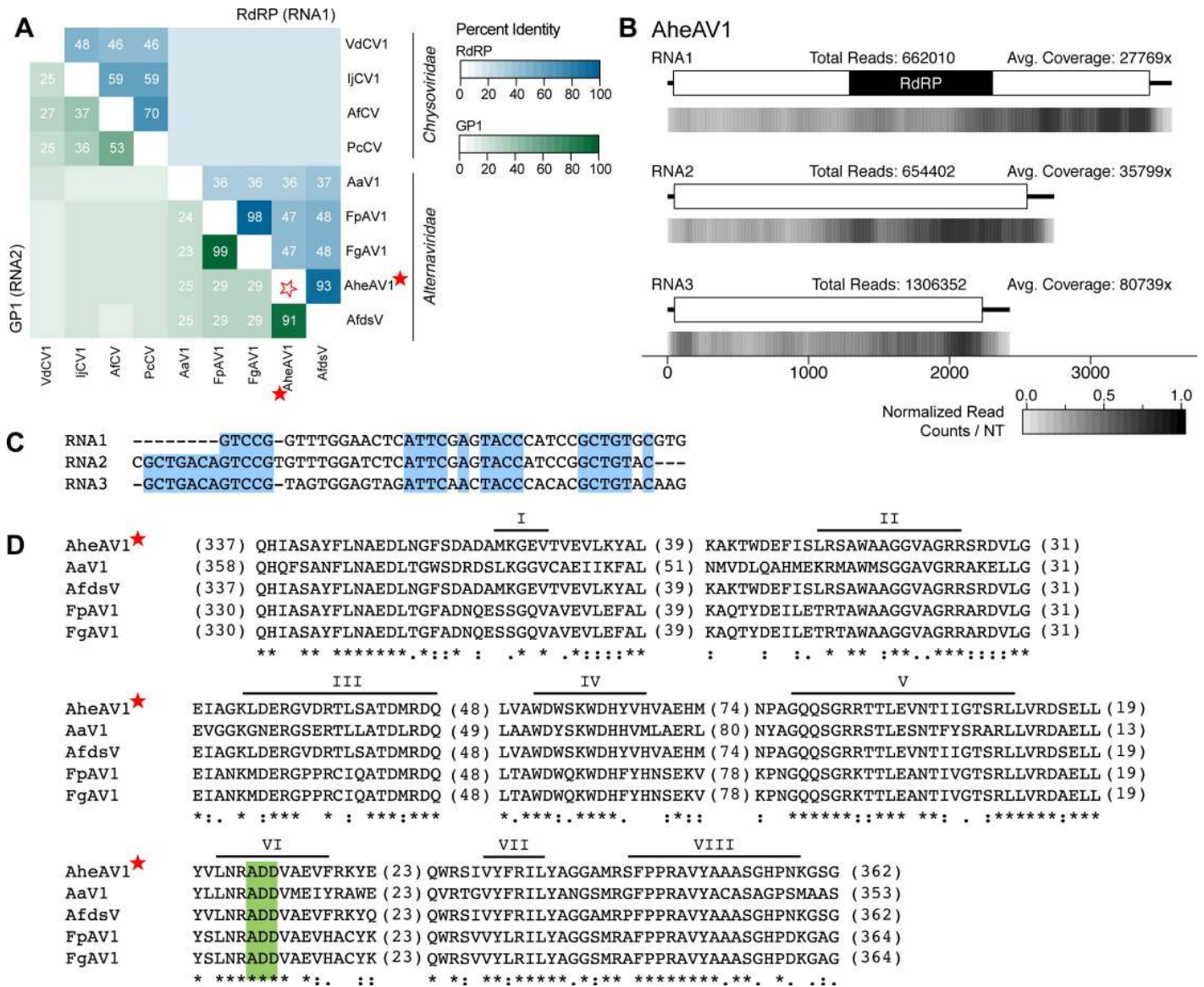


Fig 4. Analyses of *Aspergillus heteromorphus alternavirus 1* (AheAV1), a member of the recently proposed family *Alternaviridae*. (A) Percent identity matrix, generated by Clustal-Omega 2.1, of alternaviruses and members of the *Chrysoviriidae* family. The top half of the matrix, in the blue scale, is the percent identity of the RdRP protein coding sequence (from the RNA1 segment) and the bottom half of the matrix, in the green scale, is the percent identity of the second largest genomic fragment, encoding GP1. For sake of clarity, the 100% identity along the diagonal has been removed and outlined stars are used to note junction point for the row and column of the virus identified. Values ≥ 20% identity are noted. (B) Density plot of read counts per nucleotide across the viral genomic segments of AheAV1. Graphical depiction of the genome structure, predicted ORFs and motifs for each segment are on top and a heatmap of normalized read counts is below each diagram. Read counts are normalized to a 0 – 1 scale, where the maximum read count for a given genome equals 1. Total reads mapped to the virus genome and average read coverage are noted. (C) Alignment of 5'UTR sequences of the three genomic segments of AheAV1. Conserved nucleotides are highlighted in blue. (D) Alignment of RdRP domains of *Alternaviruses* found in (A). The eight conserved RdRP motifs of dsRNA polymerases are noted along the top of the alignment and the ADD triad, found only in *Alternavirus* sequences, is highlighted in green. The number of amino acids not shown in alignment are noted in parentheses. Viruses identified in this study are indicated with a red star. RdRP: RNA-dependent RNA Polymerase; GP1: Gene Product 1.

<https://doi.org/10.1371/journal.pone.0219207.g004>

a frameshift site are (1) a heptanucleotide sequence ‘slippery site’ with the sequence X XXY YYZ and (2) a downstream RNA pseudoknot [42]. As expected, the two open reading frames in HwTV1 are frame 3 (CP) and frame 2 (RdRP). A canonical slippery site, G GGU UUU, is present upstream of the ORF1 stop codon at position 1,940 – 1,946 nt, and two predicted pseudoknots are present immediately 3’ of the slippery site (S1B Fig). A conserved domain search

within the encoded proteins reveals the presence of the major coat protein (L-A virus) domain in ORF1, in addition to the RdRP domain in ORF2 (S1C Fig). A total of 12,385 reads align to the HwTV1 genome, resulting in an average coverage of 341x, with slightly higher coverage observed across the 5' half of the genome (S1C Fig).

Victorivirus. The remaining nine viruses in the *Totiviridae* family were found within two branches of the *Victorivirus* genus. The prototype virus of this genus, *Helminthosporium victoriae virus 190S* (HvV190S) [43], is present within one clade along with *Colletotrichum eremochloae totivirus 1* (CeTV1) (PRJNA262412; unpublished), *Colletotrichum navitas totivirus 1* (CnTV1) (PRJNA262225; unpublished), and *Penicillium digitatum totivirus 2* (PdTV2) (PRJNA254400; [23]). A second clade contains *Colletotrichum caudatum totivirus 1* (CcTV1) (PRJNA262373; unpublished), *Colletotrichum zoysiae totivirus 1* (CzTV1) (PRJNA262216; unpublished), *Penicillium digitatum totivirus 1-KH8* (PdTV1-KH8) (PRJNA254400; [23]), *Thelebolus microsporus totivirus 1* (TmTV1) (PRJNA372886; unpublished), *Tolypocladium ophioglossoides totivirus 1* (ToTV1) (PRJNA292830; [26]), and *Aspergillus homomorphus totivirus 1* (AhoTV1) (PRJNA250984; unpublished) along with other victoriviruses (Fig 2).

Alignment of the RNA-seq reads to the viral genomes reveals a wide range of average coverage values: from 8x to 526x (S2C and S2D Fig). A peak in coverage observed in CcTV1 and CzTV1 immediately precedes an octanucleotide sequence (AGGGUUC) which has been observed in the 5'UTR of some other victoriviruses including *Alternaria arborescens victorivirus 1* (AaVV1) [44] and *Epichoë festucae virus 1* (EfV1) [45]. The peak in coverage of CnTV1 immediately precedes an octanucleotide sequence slightly different than previously reported (CGGCCUUC). A similar peak in coverage was not observed for AhoTV1, which also contains the AAGGGUUC octamer sequence within the 5' UTR, although this viral sequence had the lowest average coverage. The genome structure diagrams drawn above the coverage heat maps highlight features of the viruses that are further described below.

Unlike the frame-shifting observed in totiviruses, the RdRP of Victoriviruses is expressed as a separate protein from the coat protein [46] via a termination-reinitiation strategy [47]. The AUG start codon either overlaps the stop codon of the upstream of ORF1 or slightly precedes it, and is nearly always found in the -1 frame relative to ORF1 [43]. The tetranucleotide sequence, **AUGA**, is observed for PdTV2, PdTV3, CzTV1, and AhoTV1. For CnTV1, TmTV1, and ToTV1, the last nucleotide of the ORF1 stop codon overlaps the ORF2 start codon, creating the pentamer sequences **UGAUG**, **UAAUG**, and **UAAUG** respectively. In each of these seven instances, there is the expected -1 frameshift between ORF1 and ORF2. For CeTV1 and CcTV1, a two nucleotide spacer is observed between the start codon of ORF2 and the stop codon of ORF1 (**AUGCAUGA** and **AUGAAUAA** respectively) that is similar to the sequences observed in SsRV1 and SsRV2 (**AUGAAUAA** and **AUGAGUAA** respectively) [48]. Thus ORF2 of CeTV1 and CcTV1, as well as SsRV1 and SsRV2, is found in the +1 frame relative to ORF1.

A characteristic of the coat protein sequences of victoriviruses is an abundance of alanine, glycine, and proline (A,G,P) residues, particularly towards the C-terminal end. An analysis of between the CP protein sequences from victorivirus species and totivirus species illustrates a generally elevated percentage of these three amino acids across the entire CP coding sequence of victoriviruses in comparison to totiviruses, with a notable increase observed in the two bins closest to the C-terminus of the protein (S2B Fig). The same pattern is also observed for all nine proposed victorivirus sequences identified in this study.

These analyses demonstrate that these nine sequences share conserved characteristics found in members of the genus *Victorivirus*. Criteria for species demarcation within this genus are fungal host species and percent identity of the RdRP and CP amino acid sequences [39], with a cutoff of 60% identity for both. The CP and RdRP sequences of CnTV1, TmTV1, and

ToTV1 satisfy both the sequence-based criteria of < 60% identity and unique fungal host criteria to be characterized as new species (S2A Fig).

For a growing number of recently published sequences with strong phylogenetic ties to victoriviruses, the cutoff of 60% identity for the RdRP and/or CP is not satisfied, only the criteria of a differing fungal host (S2A Fig). For five of the six other Victoriviruses identified in this study, a 60% pairwise identity would not distinguish between viral sequences, only the host-based criteria. The sixth virus, PdTV1-KH8, shares 97% (RdRP) and 98% (CP) identity with the already published sequence for *Penicillium digitatum virus 1* (PdV1) [49] thus neither the sequence-based nor host-based criteria are met. Notably, PdTV2 and PdTV1-KH8 were both identified from the same data sample, however they share only 33.5% identity. The coincident infection of *P. digitatum* KH8 by two victoriviruses is not unprecedented: *Sphaeropsis sapinea RNA virus 1* and *Sphaeropsis sapinea RNA virus 2* were isolated from the same host and share less than 60% similarity [48].

Clearly a precise sequence-based cut-off to distinguish between mycoviral species has not yet been identified. Until further biological studies to query the host range of these similar viruses can be performed, we would recommend more stringent values for pairwise percent identities for species demarcation within the *Victorivirus* genus: 80% for CP sequences and 90% for RdRP sequences in addition to the identity of the fungal host. Following this recommendation, the eight viruses, CnTV1, TdTV2, CeTV1, CcTV1, CzTV1, AhoTV1, TmTV1, and ToTV1, would be identified as new viral species within the *Victorivirus* genus. To our knowledge, this is the first mycovirus described for *C. navitas*, *C. eremochloae*, *C. zoysiae*, *C. caudatum*, *Thelebolus microsporus*, and *Tolypocladium ophioglossoides*.

Partitiviridae. Five genera have been assigned to the family *Partitiviridae*, along with fifteen viruses unassigned to a genus [50]. Genome structure of this family requires two distinct genome segments, each containing a single ORF: dsRNA1 and dsRNA2. The RdRP is found on dsRNA1 while the CP is found on dsRNA2, and highly conserved sequences can be found in the 5'UTR of both viral genome segments [50]. To date, viruses isolated from fungi have been classified into one of three of the *Partitiviridae* genera: *Alphapartitivirus*, *Betapartitivirus*, and *Gammapartitivirus*. In this study, sixteen viruses were identified that clearly group among the known partitiviruses (Fig 2).

Alphapartitivirus. Five viruses identified in this study phylogenetically group within the *Alphapartitivirus* genus: *Clohesyomyces aquaticus partitivirus 1* (CaPV1) (PRJNA372822; [11]), *Grosmannia clavigera partitivirus 1* (GcPV1) (PRJNA184372; [17]), *Drechslerella stenobrocha partitivirus 1* (DsPV1) (PRJNA236481 [13]), *Gaeumannomyces tritici partitivirus 1* and 2 (GtPV1, GtPV2) (PRJNA268052; [16]). To our knowledge, the first three are the first viruses identified in these fungal species while the two partitiviruses in *G. tritici* are the first partitiviruses from this fungus. As with other members of this genus, the genomic segment encoding the RdRP is approximately 1.9 to 2.0 kb while the genomic segment encoding the CP is between 1.7 and 1.9 kb (Table 1). Alignment of the RNA-seq reads to the viral genome sequences reveals average coverage per nucleotide ranging from 48x to 28,101x (S3B Fig). In general, the two dsRNA segments do not appear to have similar coverage per nucleotide across the segments, and in all but one instance, the dsRNA2 segment has the higher average coverage. For the two viruses isolated from *G. tritici*, 23x more reads align to GtPV1 than GtPV2 (S3B Fig).

Multiple sequence alignments of the 5'UTR regions of the two RNA genomic sequences for each identified viral genome revealed highly conserved stretches of nucleotides specific to each virus (S3C Fig). As dsRNA satellite sequences have been previously described for the alphapartitivirus *Cherry chlorotic rusty spot associated partitivirus* [51] the conserved sequences present

in dsRNA1 and dsRNA2 were used to query the remaining Trinity contigs, however no putative satellite RNAs were identified.

The criteria for species demarcation within the genus *Alphapartitivirus* are $\leq 90\%$ identity in the RdRP protein sequence and $\leq 80\%$ identity in the CP protein sequence. All five viral genomes identified in this study satisfy these criteria and as such are new species of the genus *Alphapartitivirus* (S3A Fig).

Betapartitivirus. Two virus genomes were identified as members of the genus *Betapartitivirus*: *Trichoderma citrinoviride partitivirus 1* (TcPV1) (PRJNA304029; [28]) and *Fusarium poae partitivirus 1-2516* (FpPV1-2516) (PRJNA319914; [15]). To our knowledge, this is the first partitivirus identified from *T. citrinoviride* while FpPV1-2516 is the third instance of a species of betapartitivirus described to date from the fungus *F. poae*.

The RdRP coding sequence from FpPV1-2516 shares 99.9% and 97.9% identity with FpV1 (*F. poae* strain A11) and FpV1 (*F. poae* strain 240374) respectively (S4A Fig). Similarly, the CP sequence of FpPV1-2516 shares 99.8% identity with both of the FpV1-CP sequences. Despite the high sequence similarity observed, these three viruses appear to originate from different locales. FpV1-240374 was published in 2016 following deep RNA-sequencing of the fungal strain *F. poae* MAFF 240374, a fungus isolated from a *Triticum aestivum* spikelet in Japan in 1991 [52]. FpV1 (strain A11) was identified in 1998 from the strain *F. poae* A-11 isolated from wheat grain in Hungary [53]. Finally, FpPV1-2516, identified in this study, was found in the RNA-seq data from *F. poae* strain 2516, a fungus isolated in Belgium from wheat [15]. While FpPV1-2516 is not a unique viral species, it does have a unique genome sequence and unique history from the other two known viruses and as such, has been deposited at Genbank (Table 1). A pairwise percent identity analysis of TcPV1 with other betapartitiviruses reveals that this is a unique sequence since identity values for both the RdRP and CP sequences are below the established thresholds for species demarcation of 90% identity and 80% respectively (S4A Fig) [50].

A sequence alignment of the 5'UTR sequences of the two genomic segments of TcPV1 reveals stretches of highly conserved nucleotides, characteristic of partitiviral genomes (S4C Fig). Additionally, the two genomic segments are both 2.3 kb (Table 1), which is within the described range for betapartitiviruses of 2.1 to 2.4 kb [50]. As observed with the majority of the alphapartitiviruses described above, the average coverage per nucleotide for the genomic segments of FpPV1-2516 and TcPV1 was higher for the dsRNA2 segment that encodes the CP than the RdRP segment (S4B Fig).

Gammapartitivirus. Five viruses identified in this analysis grouped with members of the genus *Gammapartitivirus* (Fig 2): *Penicillium digitatum partitivirus 1-HSF6* (PdPV1-HSF6) (PRJNA352307; unpublished) and *Pseudogymnoascus destructans partitivirus 1-20631* (PdPV1-20631) (PRJNA66121; unpublished) along with *Magnaporthe grisea partitivirus 1* (MgPV1) (PRJNA269089; [19]), *Thelebolus microsporus partitivirus 1* (TmPV1) (PRJNA372886; unpublished), and *Phyllosticta citriasiana partitivirus 1* (PcPV1) (PRJNA250394; unpublished). The latter three viruses are, to our knowledge, the first mycoviruses described for these fungal species and, as described below, satisfy the criteria to be identified as new viral species.

The dsRNA1 genomic segment encoding the RdRP for all five viruses is approximately 1.7 kb while the dsRNA2 genomic segment encoding the CP is between 1.5 and 1.6 kb (Table 1) both within the expected range of other known members of the genus *Gammapartitivirus* of 1.6 to 1.8 kb for dsRNA1 and 1.4 to 1.6 kb for dsRNA2 [50]. A multiple sequence alignment analysis of the 5'UTR sequences from the two genomic segments of each virus revealed stretches of conserved nucleotides within four of the five viruses; PcPV1 was the exception, however the 5'UTR sequence of dsRNA1 is only 17 nt (S5C Fig). Attempts to extend this sequence by manually querying the RNA-seq data were unsuccessful.

Total reads aligned and average coverage per nucleotide were calculated following alignment of the RNA-seq reads to the viral genome sequences. In general, the average coverage per nucleotide for dsRNA1 is similar to that of dsRNA2 (S5B Fig), a departure from the alignment results for the alphapartitiviruses and betapartitiviruses described above.

A pairwise percent identity analysis compared the RdRP coding sequences and the CP coding sequences for the five viruses to other sequences within the genus *Gammapartitivirus* (S5A Fig). The current criteria for species demarcation is 90% for the RdRP sequence and 80% for the CP sequence [50]. Genic and protein sequences from PdPV1-20631 share 100% identity with the virus PdPVpa from *Pseudogymnoascus destructans* 80251, although there are differences observed in the length of the 5' and 3' UTR sequences. The RdRP and CP protein sequences of PdPV1-HSF6, isolated from *Penicillium digitatum*, shares 95% identity with the proteins from *Penicillium stoloniferum virus S* (PsVS). Due to the high sequence similarity and the relative close relationship between the two fungal hosts from the *Penicillium* genus, this virus sequence does not meet the criteria to qualify as a new viral species. Nonetheless, these viral sequences have been deposited in GenBank due to the differences within the nucleotide sequences from published viral genomes and the identity of the host fungal strain. The RdRP and CP protein sequences from MgPV1, PcPV1, and TmPV1 do not share significant identity with known gammapartitivirus sequences and are isolated from new fungal hosts. As such, these three viruses are characterized as new viral species, belonging to the genus *Gammapartitivirus*.

New genus: *Epsilonpartitivirus*. Two viral genomes, *Colletotrichum eremochloae partiti-virus 1* (CePV1) (PRJNA262412; unpublished) and *Penicillium aurantiogriseum partiti-like virus* (PaPIV) (PRJNA369604; [12]), were identified that are phylogenetically related to an emerging group of partitiviruses that are separate from, but related to, the genus *Gammapartitivirus* (Fig 2); this group has been tentatively named *Epsilonpartitivirus* [12]. Identification of the PaPIV sequences was expected as this sample was RNA-seq data from *Cryphonectria parasitica* infected with PaPIV, previously isolated from *P. aurantiogriseum* by the same researchers [12]. To distinguish this viral genome from that isolated from the original host, the virus has been designated PaPV1-cp. Also, the slight modification of the virus name, from partiti-like to partitivirus, reflects a recent improved understanding that this is a partitivirus due to the identification of a RNA segment encoding a putative CP [12].

Members of this proposed genus have been identified from both fungal and invertebrate hosts, although currently the viral sequences form two sister clades separated by kingdom (Fig 2). Genome sizes of the dsRNA segments of putative epsilonpartitiviruses are from 1.7 to 1.9 kb for dsRNA1 and 1.5 to 1.8 kb for dsRNA2 (Table 1). Using a pairwise comparison analysis, a value of 85% identity for the RdRP protein sequence and 75% for the CP sequence would currently be appropriately stringent for species demarcation of viruses isolated from different fungal hosts (Fig 5A).

Observed here, sequences isolated from the PaPIV1-infected *C. parasitica* sequencing sample contain a number of SNPs in comparison to the source PaPIV sequence; these changes may reflect an adaptive response by the virus to a new fungal host [12]. A positive effect on fungal growth is observed whereby the virus confers a positive effect on fungal growth of *C. parasitica* in the presence of 2% salts that has not yet been confirmed in the original host *P. aurantiogriseum*. Read coverage along the viral genome segments isolated from *C. parasitica* is on average low, 18x for dsRNA1 and 5x for dsRNA2 (Fig 5B), similar to the published report [12].

Both of the viral genome segments of CePV1 share the greatest percent identity with PaPIV1, but below the threshold suggested above, at 84% and 73% for the RdRP and CP respectively (Fig 5A). Similar to other partitiviruses, a multiple sequence alignment of the

5'UTR sequences of the two genome segments of CePV1 reveals regions of sequence conservation (Fig 5C).

Initially a mycoviral sequence was isolated from a *Colletotrichum navitas* sequencing sample (PRJNA262223; unpublished); however, other than small variances within the 5' and 3' UTRs, the nucleotide sequence is identical to that of CePV1. It is possible that two different, closely related, fungi are infected with the same virus, however as noted above with respect to the PaPIV1-infected *C. parasitica*, SNPs can quickly appear within a viral genome sequence when in a new host. Therefore further analyses were performed to determine the likelihood of a CePV1-like sequence hosted by *C. navitas*. First it was determined that, after controlling for the unaligned reads pool size, 2.6x more viral reads were present in the *C. eremochloae* sample than the *C. navitas* sample. Second, to identify possible cross-contamination between the two species within the sequencing data, non-virus-derived reads were reciprocally aligned to the other fungal genome. Thus reads from *C. navitas* were aligned to the *C. eremochloae* genome, and vice versa. Here we observed that 5% of reads from *C. eremochloae* aligned to the *C. navitas* genome while 65% of reads in the *C. navitas* pool aligned to the *C. eremochloae* genome. Evaluating both of these results, we propose that the likely host of this viral sequence is *C. eremochloae* and the identification of the same sequence within the *C. navitas* sample is likely due to cross contamination from the *C. eremochloae* sample. Despite the possible ambiguity in fungal host for this viral sequence, we recommend its inclusion in the recently suggested genus *Epsilonpartitivirus* as a new viral species.

New genus: *Zetapartitivirus*. The final two viruses identified within the *Partitiviridae* branch of the phylogenetic tree (Fig 2) are *Penicillium brasilianum partitivirus 1* (PbPV1) (PRJEB7514; [22]) and *Delitschia confertaspora partitivirus 1* (DcPV1) (PRJNA250740; unpublished). Both viruses, the first to be described in these fungal species, phylogenetically group with *Alternaria alternata partitivirus 1* (AaPV1) [54] and *Botryosphaeria dothidea partitivirus 1* (BdPV1) [55], a distinct clade within the *Partitiviridae* that is related to, but separate from, the genus *Gammapartitivirus* (Fig 2).

These four viruses share many sequence-based similarities. A 13-nt highly conserved sequence (TCACAAAYATYAYA) is present in the 5'UTR sequence of both dsRNA genome segments from all four viruses (Fig 5F). Additionally, a 10-nt conserved sequence (TYTCAT-TARR) is found within the 3'UTR sequence of both dsRNA segments of all four viruses (Fig 5G). While these four viruses contain the six conserved motifs common to RdRP proteins from other members of the *Partitiviridae*, including the GDD triad in motif VI, there are distinct differences that are shared that distinguish these viruses from the existing genera (Fig 5D). Further, the range of lengths for the RdRP protein, from 569 to 572 aa, and CP protein, from 453 to 501 aa (Table 1), are unique ranges among the currently described genera [50]. A pairwise percent identity analysis of the RdRP and CP sequences from the four viruses suggests a threshold for species demarcation at 90% for the RdRP sequence and 80% for the CP sequence, similar to other genera within the family (Fig 5A).

Alignment of the RNA-seq reads to the viral genomes demonstrates that dsRNA2 has a higher average coverage per nucleotide for both viruses, with approximately 2x more coverage observed between the segments of PbPV1 and 7x between the segments of DcPV1 (Fig 5E).

We would propose that DcPV1 and PbPV1, along with AaPV1 and BdPV1 belong to a new genus within the family *Partitivirus*, called *Zetapartitivirus*. The genomic segment and protein sequence lengths are consistent and unique within the *Partitivirus* family, and the RdRP protein sequences form a distinct clade that is separate from the closest genus, *Gammapartitivirus* (Fig 2).

***Chrysoviriidae*.** To date, nine species have been classified as belonging to the Chrysoviriidae family of dsRNA viruses that infect either ascomycota or basidiomycota fungi [56].

row and column of the viruses identified. Two cutoff values were applied and as such only percent identities $\geq 29\%$ among RdRP sequences and $\geq 17\%$ among CP sequences are indicated. Virus names in italics refer to viruses isolated from insect, not fungal, hosts. Bolded virus names are member species recognized by ICTV. (B and E) Density plot of read counts per nucleotide across the viral genome segments for the two epsilonpartitiviruses (B) and two zetapartitiviruses (E). Graphical depiction of the genome structure, predicted ORFs, and RdRP motif are along the top and a heatmap of normalized read counts is below. Read counts are normalized to a 0–1 scale, where the maximum read count for a given genome equals 1; heatmap scale found in (E) applies also to (B). Total reads mapped to the virus genome and average read coverage are noted for each virus. (C) Alignment of the 5' UTR sequences from the two genomic segments of the new species of epsilonpartitivirus, CePV1. Conserved nucleotides are highlighted in blue and numbers in parentheses indicate nucleotides not shown. (D) Alignment of the six conserved motifs within the RdRP protein sequence of all four proposed zetapartitiviruses. Amino acids highlighted in blue are only common among zetapartitiviruses while residues in grey are shared with gammapartitiviruses. (F and G) The 5'UTR (F) and 3'UTR (G) sequences from both genomic segments for all four zetapartitiviruses were aligned using MAFFT and a graphical representation of the most highly conserved sequences found is depicted as a sequence logo, generated using WebLogo version 3, whereby the height of a nucleotide indicates the sequence conservation at that position (<http://weblogo.threeplusone.com/>). A consensus sequence for each UTR is indicated below in grey. Red stars indicate viruses identified in this study. PV: partitivirus; RdRP: RNA-dependent RNA Polymerase; CP: Coat Protein.

<https://doi.org/10.1371/journal.pone.0219207.g005>

Members of this family have multipartite genomes, consisting of three to five segments. Here we identify two viral genomes with sequence similarity to Chrysovirus: *Beauveria bassiana chrysovirus 1* (BbCV1) (PRJNA306902; [10]) and *Penicillium raistrickii chrysovirus 1* (PrCV1) (PRJNA250734; unpublished) (Fig 2).

Phylogenetically, chrysovirus are grouped into two distinct clusters, where cluster 1 contains members with either three or four segmented genomes, while cluster 2 contains viruses with either four or five segments [56]. The two viral sequences identified in this study, BbCV1 and PrCV1, belong with cluster 1 chrysovirus; both genomes are comprised of four genomic segments and the RdRP sequences phylogenetically group with the ICTV-recognized members within cluster 1, *Isaria javanica chrysovirus 1* (IjCV1) [57] and *Penicillium chrysogenum virus* (PcCV1) [58] (Fig 2). Further, as detailed below, we recommend that these are both new viral species within the genus *Chrysovirus*.

To our knowledge, these are the first chrysovirus identified for both of these fungal hosts, and indeed the first virus from *P. raistrickii* described, thus satisfying the host-based criteria for species demarcation. Following identification of the RdRP-encoding contig, the identity of the other three segments was determined via a BLASTP search using protein sequences of known chrysovirus against the predicted protein sequences of the Trinity contigs. The combined length of dsRNAs for both BbCV1 and PrCV1 are 12.46 and 12.56 kb respectively (Table 1), within the observed range of other chrysovirus (11.5 to 12.8 kb) [56]. In general, chrysovirus genome segments are numbered by size, however segment 4 of PrCV1 is slightly longer than segment 3 (Table 1), similar to segment 4 of *Amasya cherry disease associated chrysovirus* (AcdaCV) which is also larger than its respective segment 3 [59]. A BLASTP analysis confirms that segment 4 of PrCV1 best matches the segment 4 of *Penicillium chrysogenum virus* and thus is numbered accordingly. An alignment of the RNA-seq reads to the viral genome segments reveals uniform average coverage per nucleotide for the four genomic segment of PrCV1 (from 15x to 32x) and slightly more variable average coverage of BbCV1 segments, ranging from 10x to 72x (S6D Fig). The dsRNA genomic segment lengths therefore also satisfy the criteria for species demarcation within the genus *Chrysovirus*.

The 5'UTR of chrysovirus is generally between 140 and 400 nt in length and contains two regions of high sequence conservation, Box 1 and Box 2 [56], although a number of recently characterized viruses have shorter 5'UTRs, including *Cryphonectria nitschkei chrysovirus 1* (CnCV1) which has 82 nt 5'UTR for segment 1 [60] and AcdaCV which has 5'UTR lengths of 87, 96, 95, and 106 bp [59]. With the exception of the 5'UTR of BbCV1-segment 1 (106 nt), the 5'UTR of all segments discovered for BbCV1 and PrCV1 fall within the published range (S6B and S6C Fig). A highly conserved region is present within the area corresponding to “Box 1” within the 5'UTR of the four genome segments of BbCV1 (S6B Fig, top panel) and the expected “CAA” sequence is also observed in all sequences (S6B Fig, bottom panel). Similarly both conserved regions are observed within the 5'UTR of PrCV1 (S6C Fig). Thus the size and

composition of the 5'UTR sequences of BbCV1 and PrCV1 also satisfy the criteria for chrysovirus classification.

The final criteria considered for species demarcation is the percent identity of the RdRP and CP protein sequences relative to other members within the genus, where the thresholds are 70% and 53% respectively [56]. BbCV1 RdRP and CP share 83% and 69% identity with the related proteins of IjCV1, while PrCV1 RdRP and CP protein sequences share 88% and 79% identity with those of *Penicillium chrysogenum virus* (PcCV1) (S6A Fig). Since BbCV1 was isolated from a new fungal host, we propose classifying it as a new species of chrysovirus, while the relative relatedness between the hosts *P. raistrickii* and *P. chrysogenum* suggest that PrCV1 is not a new species of chrysovirus. Further identification of additional viruses plus biological characterization of host specificity will provide the needed insights to properly classify viruses within this family.

Endornaviridae. Members of the *Endornaviridae* family of viruses have long, single segment, dsRNA genomes, generally encoding a single polyprotein. In addition to the RdRP domain and helicase domain (Viral helicase 1 or DEADx helicase) found in all endornaviruses, viral genomes may also encode a viral methyltransferase domain and/or a glycosyltransferase domain. Recently a two-genus split was proposed [61] where the common characteristics of clade I (*Alphaendornavirus*) include a plant-host, larger genomes (ranging from 13.8 to 17.5 kb), the presence of a UGT domain and a site-specific nick towards the 5' end of the genome sequence [62], while members of clade II (*Betaendornavirus*) have fungal-hosts and generally have genome sizes ranging from 9.5 to 11.4 kb, encode a viral methyltransferase domain but not a UGT domain, and lack the site-specific nick [61].

Here four new virus sequences were identified that group with known endornaviruses (Fig 6A), *Gyromitra esculenta endornavirus 1* (GeEV1) (PRJNA372840; unpublished) and *Morchella importuna endornavirus 1, 2, and 3* (MiEV1, MiEV2, MiEV3) (PRJNA372858; unpublished). These are the first viruses, to our knowledge, identified from both *G. esculenta* and the morel mushroom, *M. importuna*.

Alphaendornavirus. A phylogenetic analysis of the RdRP motif from related endornaviruses and the viruses identified in this study reveals MiEV2 and GeEV1 form a well-supported clade with other fungal endornaviruses, including *Ceratobasidium endornavirus A* (CEVA) [63], and *Rhizoctonia solani endornavirus 2* (RsEV2) [64], that is related to, but separate from alphaendornaviruses isolated from plant hosts, including the type species *Oryza sativa endornavirus* (OsEV) [65] (Fig 6A). The genome sizes of MiEV2 and GeEV1 are 15.3 and 14.5 kb respectively, well within the observed range of other *Alphaendornaviruses* (Table 1). Additionally, a conserved domain search within the protein coding sequences determined both viruses encode a viral helicase and an RdRP domain, but not a viral methyltransferase domain, nor a UGT domain (S7B Fig). This combination of protein domains appears to be shared among mycoviruses and not plant-hosted endornaviruses.

GeEV1 has a predicted AUG at position 5, resulting in a shorter 5' UTR sequence than observed in other endornaviruses. The predicted start codon for MiEV2 however is at position 476, resulting in an unexpectedly long 5'UTR; as the entire sequence upstream of this codon is in-frame with the coding sequence, it is possible that this genome sequence is partially truncated and is missing the true AUG sequence. As manual analysis of the RNA-seq data did not result in an extension of the 5' region of MiEV2, we recommend further analysis of the biological material to confirm the 5' region of this virus. An alignment of RNA-seq reads to the two viral genomes reveals uniform coverage across both genome sequences, with a total of 4977 reads and 8171 reads respectively (S7B Fig).

The two criteria for species demarcation within the genus *Alphaendornavirus* are difference in host and an overall nucleotide sequence identity below 75%. Both viruses described here

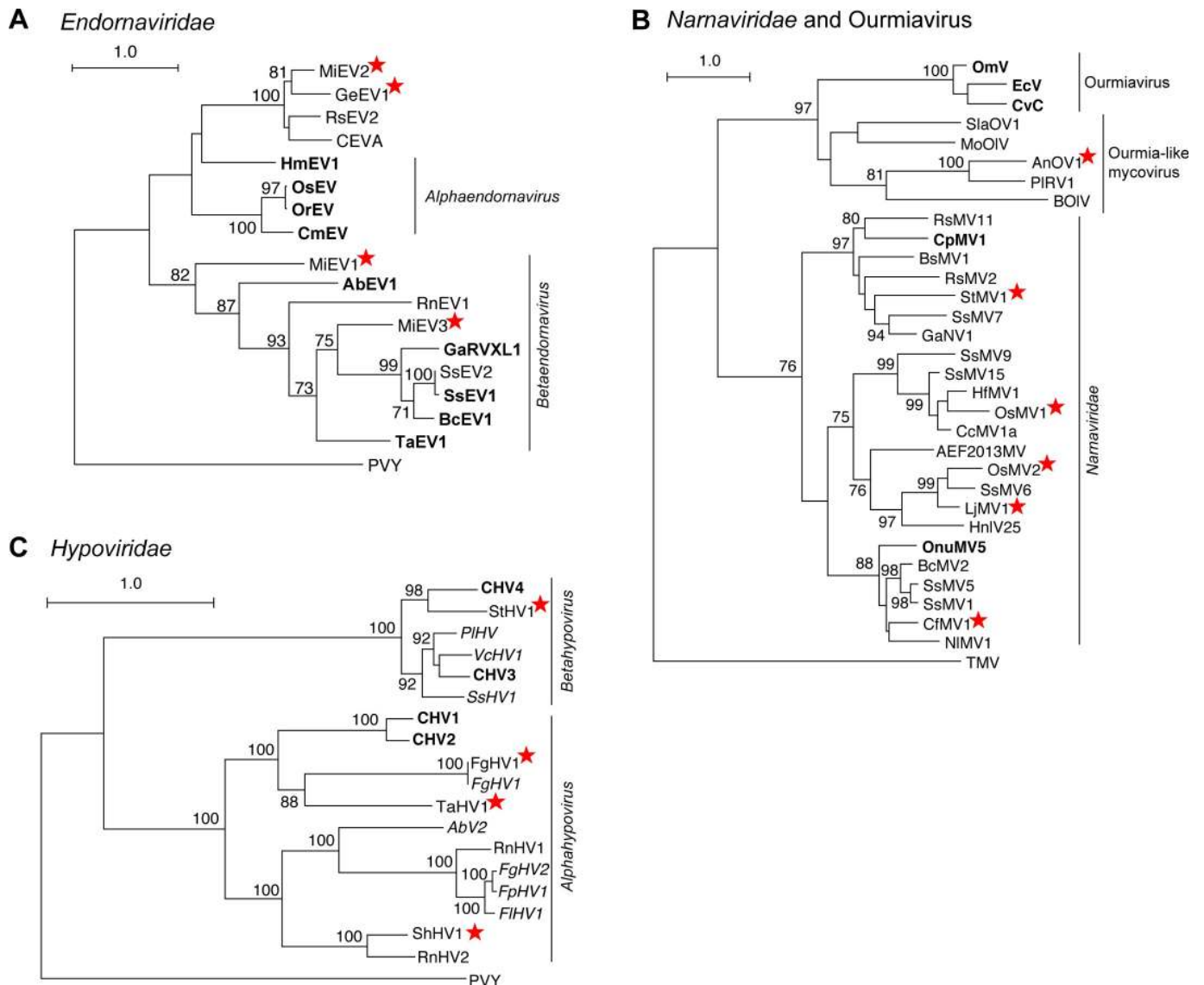


Fig 6. Phylograms for three different families of (+)ssRNA viruses. RdRP motifs from the *Endornaviridae* family (A), *Namnaviridae* and *Ourmiaviruses* (B), and *Hypoviridae* (C) family were used to construct maximum likelihood phylogenetic trees generated with RaxML, using the LG+G+I amino acid substitution model. Full names of viruses and sequences can be found in [S2 Table](#), including the outgroup, *Potato Virus Y* (PVY). Bolded names are ICTV-recognized member species while italicized names are related, but unclassified viruses. Full names and accession numbers for protein sequences used in the alignment are in [S2 Table](#). Scale bar on the phylogenetic tree represents 1.0 amino acid substitutions per site and numbers at the nodes indicate bootstrap support over 70% (1000 replicates). Viruses identified in this study are noted with a red star.

<https://doi.org/10.1371/journal.pone.0219207.g006>

satisfy these criteria. Indeed, nucleotide identity among *Endornaviruses* is generally below 50%. An analysis of the protein sequences of the RdRP motif and the viral helicase motif reveal similar results, as sequences are between 39% and 67% identical for the RdRP sequences and 26% to 51% identical among the helicase sequences ([S7A Fig](#)). Together these analyses support the recognition of GeEV1 and MiEV2 as new species within the *Endornaviridae* family, and the genus *Alphaendornavirus* in particular.

Betaendornavirus. The two other endornaviruses identified were MiEV1 and MiEV3, which group with the betaendornaviruses. MiEV3, in particular, forms a strong phylogenetic relationship with other characterized betaendornaviruses, while MiEV1 is the outermost

member of this clade (Fig 6A). Unexpectedly, the genome sizes of both viruses are more than 25% larger than other betaendornaviruses; MiEV1 has a genome size of 16.5 kb while MiEV3 has a genome size of 14.3 kb (Table 1). However, the motifs within the protein sequences are as expected, as both MiEV1 and MiEV3 contain a viral methyltransferase domain, helicase domain(s), and an RdRP domain (S7D Fig). Specifically MiEV1 and MiEV3 both encode a DEADx helicase domain and a viral helicase 1 domain, similar to the type species *Sclerotinia sclerotiorum endornavirus-1* (SsEV1) (NC_021706.1). Alignment of the RNA-seq reads to the viral genomes reveals an average coverage per nucleotide of 29x and 51x for MiEV1 and MiEV3 respectively (S7D Fig), and coverage appears uniform across both genomes.

Similar to the alphaendornaviruses, nucleotide identity is generally below 50%, with the exception of SsEV1 and *Sclerotinia sclerotiorum endornavirus 2* strain DIL-1 (SsEV2) which share 82% nucleotide identity. Further analysis of the protein sequences from the RdRP motifs and viral methyltransferase motifs determined that a percent identity threshold for either motif of 75% would be appropriate for species demarcation (S7C Fig). In particular, both MiEV1 and MiEV3 are well below both a nucleotide-based or a protein-motif based threshold. Thus, despite the same fungal host, the uniqueness of the viral genomes sequence for both viruses suggests that both MiEV1 and MiEV3 are new species of the family *Endornaviridae*, and belong within the genus *Betaendornavirus*.

Positive-sense single-stranded RNA ((+)ssRNA) viruses

Tobamo-like virus. A virus sequence was identified in an *Acidomyces richmondensis* sample (PRJNA250470; unpublished) with a single segment genome sequence, 10,291 nt in length, that encoded four predicted ORFs (Table 1). The gene product from ORF2 contains the RdRP domain (Fig 7B) and a BLAST-P search of the 'nr' database with this sequence identified two related, full length mycoviruses, *Macrophomina phaseolina tobamo-like virus 1* (MpTIV1) [5] and *Podospaera prunicola tobamo-like virus 1* (PpTIV1) [66]. As such, this newly identified virus has been named *Acidomyces richmondensis tobamo-like virus 1* (ArTIV1).

For all three mycoviral sequences the first and largest ORF contains a viral methyltransferase and viral helicase domain and ORF2 encodes the RdRP domain (Fig 7B). ORF3 and ORF4 in MpTIV1 have been putatively identified as encoding a movement protein and coat protein respectively [5].

A phylogenetic analysis of the RdRP motif from the three mycoviruses compared to select members of the families *Virgaviridae*, *Bromoviridae* and genus *Idaeovirus* demonstrated that ArTIV1 is more closely related to the two previously identified mycoviruses than known plant viruses (Fig 7A). Further, the group of mycoviruses form a sister clade to the *Virgaviridae* viruses. Of the viruses currently characterized from this family, only members of the genus *Tobamovirus* have a non-segmented genome similar to ArTIV1 and relatives; other members have multipartite genomes [67].

To further characterize this emerging group of mycoviruses, a percent identity analysis was undertaken, comparing the amino acid sequences from the RdRP and viral helicase motifs to members of the *Virgaviridae* family. The RdRP motif sequences among the mycoviruses share 56 to 63% identity with one another, but less than 44% identity with *Virgaviridae* viruses (Fig 7C). Further, sequences from the viral helicase domain were 59 to 62% identical among mycoviruses, and less than 40% shared identity outside of this clade. Due to these values, it appears that all three viruses are unique viral sequences.

Finally, the RNA-seq data was aligned to the ArTIV1 genome sequence, which revealed uniform coverage along the length of the genome, averaging 1,167 reads per nucleotide (Fig 7B).

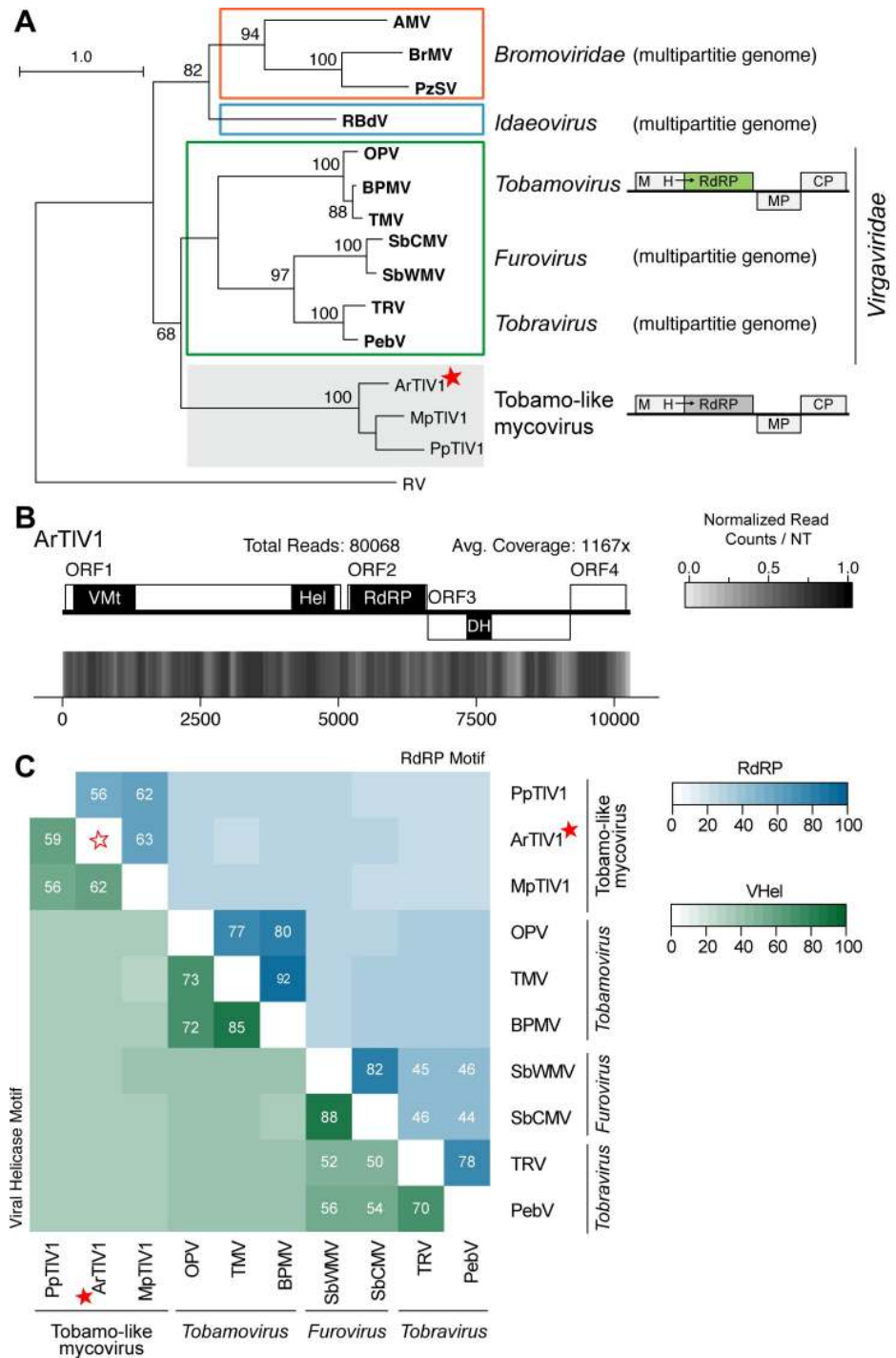


Fig 7. Analyses of *Acidomyces richmondensis tobamo-like virus 1* (ArTIV1) and related viruses. (A) Phylogram of the RdRP-motif for ArTIV1, related mycoviruses, and select members of the genus *Tobamovirus* and family *Bromoviridae*; sequence from *Rubella Virus* (RV) serves as the outgroup. Bolded names are ICTV-recognized member species. Full names and accession numbers for protein sequences used are in S2 Table. Numbers at the nodes indicate bootstrap support values $\geq 65\%$ (1000 replicates). Scale bar on the phylogenetic tree represents 1.0 amino acid substitutions per site. Diagrammatic genome structures are to the right of the phylogram. (B) Density plot of read counts per nucleotide across the ArTIV1 viral genome. Graphical depiction of the genome structure, predicted ORFs, and encoded motifs are along the top and a heatmap of normalized read counts is on the bottom. Reads are normalized to a 0–1 scale, where the maximum read count equals 1. Total reads mapped to the virus genome and average read coverage are noted. (C) Percent identity matrix, generated by Clustal-Omega 2.1. The top half of the matrix, in the blue

scale, is the percent identity of the RdRP motif sequence while and the bottom half of the matrix, in the green scale, is the percent identity of the residues from Viral Helicase motif. For sake of clarity, the 100% identity along the diagonal has been removed and outlined stars are used to note junction point for the row and column of ArTIV1. Cutoff values of 44% and 40% were applied to the RdRP and Helicase matrices respectively. Viruses identified in this study are noted with a red star. VMt: Viral Methyltransferase; VHel: Viral Helicase; RdRP: RNA-dependent RNA Polymerase; DH: DEXDc Helicase; ORF: Open Reading Frame.

<https://doi.org/10.1371/journal.pone.0219207.g007>

To date all viruses of the *Virgaviridae* family infect a plant host, although many members are transmitted by nematodes or plasmodiophorids, a group of parasitic microscopic organisms unrelated to fungi or oomycetes [67]. Tobacco mosaic virus (TMV), the best characterized *Tobamovirus*, can infect, replicate, and persist in the ascomycota fungi *Colletotrichum acutatum* [68]. Further, a virus from the *Bromoviridae* family, Cucumber mosaic virus (CMV), was shown to naturally infect the phytopathogenic fungi *Rhizoctonia solani* and as well as replicate within *Valsa mali* but not *C. parasitica*, and *F. graminearum* [69]. Due to the close contact between phytopathogenic fungi and plants a change in host range of a tobamo-like virus is plausible. Interestingly, the *A. richomondensis* strain analyzed in this study was isolated from acid mine drainage biofilms [70] and is not known to be phytopathogenic. Acid mine drainage biofilms are complex communities comprised of different fungi along with bacteria, archaea and bacteriophages [71,72]. The role of the mycovirus identified in this study in such an extreme and heterogeneous environment remains to be determined.

Ambiguiviridae. Four viral genome sequences were identified that share similarity to unclassified (+)ssRNA mycoviruses: *Trichoderma harzianum* ssRNA virus 1 (ThAV1) (PRJNA216008; [29]), *Setosphaeria turcica* ssRNA virus 1 (StAV1) (PRJNA250530; unpublished), *Verticillium longisporum* ssRNA virus 1 (VlAV1) (PRJNA308558; unpublished), and *Periconia macrospinosa* ssRNA virus 1 (PmAV1) (PRJNA262386; [24]). To our knowledge, this is the first viral sequence described for the later three fungal species, and the first ssRNA virus isolated from *T. harzianum*.

Six other viruses appear related to these four newly identified viruses: *Diaporthe RNA virus* (DRV), *Magnaporthe oryzae RNA virus* (MoRV), *Soybean leaf-associated ssRNA virus 1 and 2* (SlaRV1, SlaRV2), *Verticillium dahliae RNA virus* (VdRV), and *Sclerotinia sclerotiorum umbra-like virus 1* (SsUIV1) [5,6,73–75]. The RdRP protein sequence of this group of mycoviruses shares some sequence similarity to RdRPs of recently described (+)ssRNA viruses from insects, as well as members of the plant-infecting family *Tombusviridae*. Phylogenetic analysis demonstrates the fungal viruses form a well-supported clade that is separate from both the insect and plant viruses (Fig 8A). A pairwise identity analysis of the RdRP motif region from the viruses analyzed in the phylogenetic tree reveals a range of identities between the fungal viruses, from 36% to 63%, while only sharing 25% to 33% identity with the insect viruses and 26% to 32% identity with the members of the *Tombusviridae* (Fig 8D). A similar analysis using the ORF1 protein coding sequence, from only the mycoviral genomes, demonstrates 18 to 38% identity among sequences (Fig 8D).

Together these ten fungal viruses are characterized by a single segment genome, 2.6 kb to 4.5 kb in length, that encodes two ORFs, where the second ORF contains the RdRP domain (Table 1). The amber stop codon ‘UAG’, found at the end of the first ORF, has been suggested to result in readthrough to produce a fusion protein with the downstream RdRP [75]. All members of this group of viruses contain the amber stop codon, and both ORF1 and ORF2 are found in the same frame, a requirement for readthrough. Additionally, all 10 viruses share a number of key differences in residues commonly conserved within (+)ssRNA viruses. For instance, the second aspartic acid residue in motif A, a feature of (+)ssRNA viruses [76], is instead a glutamic acid (Fig 8C). Additionally, these 10 viruses have a GDN triad in motif C

versus the canonical GDD found in (+)ssRNA viruses (Fig 8C). This triad has previously been observed within some negative-strand (-)ssRNA viruses [76], although within (+)ssRNA viruses, modification of the GDD to GDN has repeatedly demonstrated a detrimental effect on enzymatic activity [77–79]. Finally, within motif D, a positively-charged lysine, common to nearly all RdRP sequences [76,80], is a nonpolar proline or valine residue (Fig 8C).

The 5'UTRs range in size from 283 bp to 889 bp in length for all the mycoviruses in this group, with the exception of VdRV1 and VIAV1, the two viruses isolated from species of *Verticillium*; these 5'UTRs are 11 nt and 54 nt respectively. Similarly, the 3'UTR sequences range in length from 372 bp to 1279 bp, while those of VdRV1 and VIAV1 are 182 nt and 29 nt respectively (Fig 8C). Average coverage per nucleotide across the viral genomes of the four identified viruses is fairly uniform across the genome with ranges from 15x for VIAV1 to 734x for PmAV1 (Fig 8B).

The characteristics of these four identified viruses and their closest mycoviral relatives supports the creation of a new family of viruses within the (+)ssRNA class. Initially the host of DRV was identified as *D. ambigua*, although subsequent analyses determined the host was instead *D. perijuncta* [81]; the name has been updated by ICTV, but remains uncorrected in other databases. We propose the establishment of the family *Ambiguiviridae* that contains DRV and other recently discovered viruses related to it, including the four described in this study. The proposed name references the unusual nature of the GDN triad within the RdRP motif of these (+)ssRNA viruses and the ambiguity related to the host of DRV, the first virus described for this group.

Yadokariviridae. Samples *Aspergillus homomorphus* (PRJNA250984; unpublished) and *Penicillium digitatum* (PRJNA352307; unpublished), which yielded a totivirus (AhoTV1) and partitivirus (PdPV1) respectively, each also contained a second, unrelated contig containing an RdRP motif. BLAST-X analysis of these nucleotide sequences determined they were both related to a group of unclassified, (+)ssRNA viruses, that includes four other mycoviruses. As described below, these two viruses are new members of a recently proposed family, *Yadokariviridae* [82].

Aspergillus homomorphus yadokarivirus 1 (AhoYV1) and *Penicillium digitatum yadokarivirus 1* (PdYV1), along with other members of this group, are characterized by having single segment genomes that encode a single predicted ORF, the RdRP (Table 1). Genome sizes range from 3,144 nt (PdYV1) to 5,089 (*Yado-kari virus 1*, YkV1). The length of the RdRP protein sequence is predicted to range from 962 aa for *Aspergillus foetidus slow virus 2* (AfsV2) [83] to 1,430 aa for YkV1 [84]. The lengths of the 5'UTRs are all greater than 500 nt, except for PdYV1. Attempts to extend the sequence of the 5'UTR of PdYV1 using the RNA-seq data were not successful; however, as the average coverage per nucleotide for PdYV1 was low, just 23x (Fig 9C), additional experiments with the fungal source material may be fruitful. Lengths of the 3'UTRs for all six viruses are also largely similar, between 200 to 400 nt, except for YkV1 which has a 3'UTR of 1,223 nt. A genome segment belonging to a *Yadokarivirus* that encodes a coat protein has not been identified. Alignment of the RNA-seq reads to the genomes of AhoYV1 and PdYV1 reveals an average coverage per nucleotide of 151x and 23x respectively (Fig 9B and 9C).

A phylogenetic analysis of the RdRP protein sequence reveals a well-supported clade that is most closely related to the *Caliciviridae* and separate from other known mycoviral families including *Hypoviridae* and fusariviruses (Fig 9A). AhoYV1 in particular is most closely related to AfsV2; both viruses were isolated from a member of the genus *Aspergillus* [83].

As percent identity between virus RdRP sequences is a common criteria for defining new groups of viruses, the RdRP motif sequences from the viruses in Fig 9A were analyzed to identify a potential threshold value. Indeed a threshold between 26 to 35% identity easily classifies

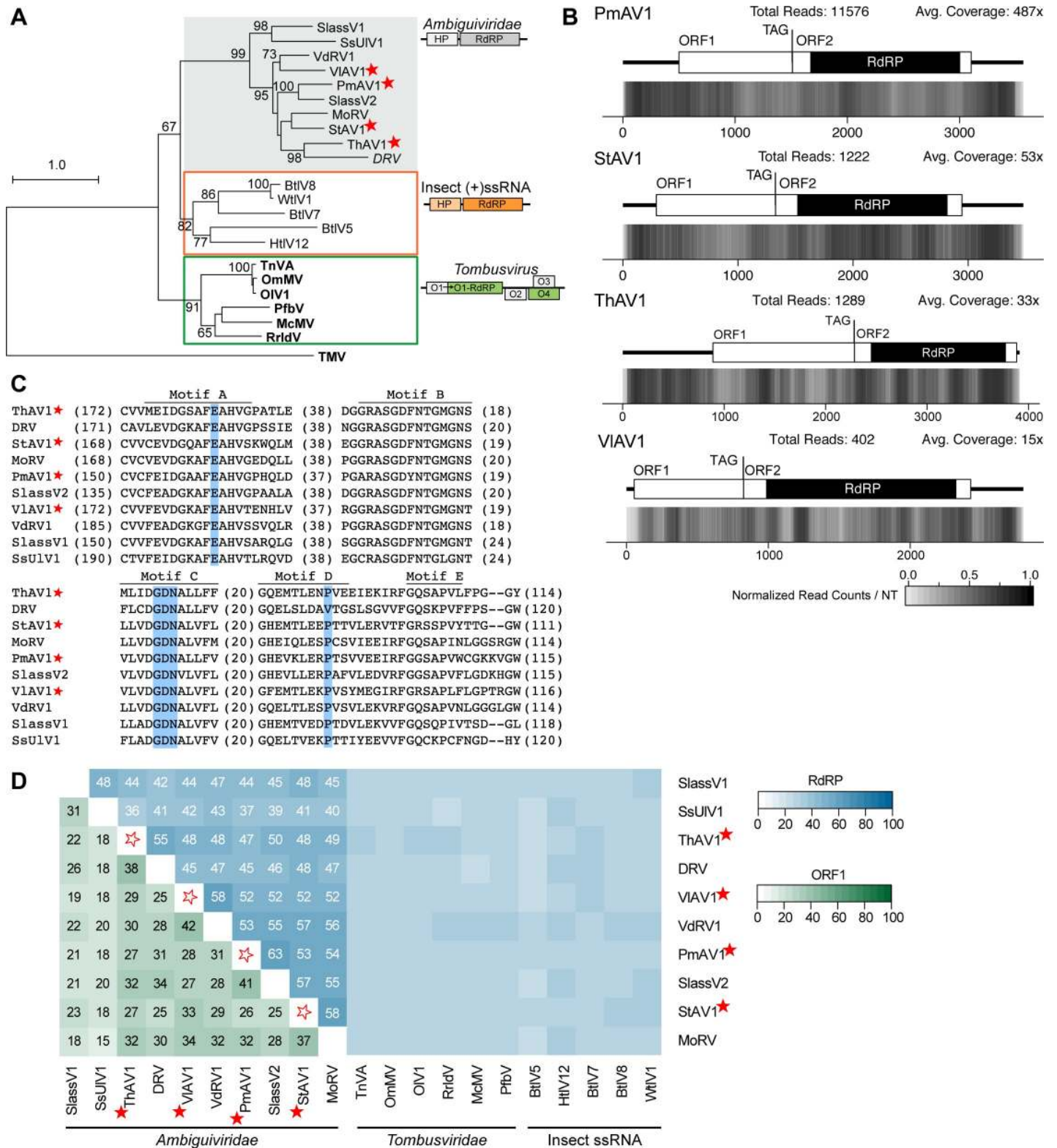


Fig 8. Analyses of four newly identified (+)ssRNA viral genomes and viruses belonging to the proposed family Ambiguiviridae. (A) Phylogram of the RdRP-motif for four new viruses and related viruses along with select members of the family *Tombusviridae* and members of an unnamed group of (+)ssRNA insect viruses; sequence from *Tobacco Mosaic Virus* (TMV) serves as the outgroup. Bolded names are ICTV-recognized member species while italicized names are recognized, but unclassified viruses. Full names and accession numbers for protein sequences used are in [S2 Table](#). Numbers at the nodes indicate bootstrap support values $\geq 60\%$ (1000

replicates). Scale bar on the phylogenetic tree represents 1.0 amino acid substitutions per site. Diagrammatic genome structures and predicted ORFs for each clade of viruses is to the right of the phylogram. (B) Density plot of read counts per nucleotide across four viral genomes. Graphical depiction of the genome structure, predicted ORFs, encoded motifs and the amber stop codon (TAG) for each genome are along the top and a heatmap of normalized read counts is on the bottom. Reads are normalized to a 0–1 scale, where the maximum read count for a given genome equals 1. Total reads mapped to the virus genome and average read coverage are noted. (C) Multiple sequence alignment of the RdRP motifs of members of the *Ambiguiviridae* family. Motifs A–E found in (+)ssRNA viral RdRP sequences are indicated along the top of the alignment. Conserved residues that differ from canonical (+)ssRNA RdRP sequences, including the unusual GDN triad found in Motif C, are highlighted in blue. Numbers in parentheses are residues not shown. (D) Percent identity matrix, generated by Clustal-Omega 2.1, of mycoviruses that comprise the proposed family *Ambiguiviridae*. The top half of the matrix, in the blue scale, is the percent identity of the RdRP motif sequence and the bottom half of the matrix, in the green scale, is the percent identity of the ORF1 amino acid sequence. For sake of clarity, the 100% identity along the diagonal has been removed and outlined stars are used to note junction point for the row and column of the viruses identified. All RdRP motif sequences share $\geq 35\%$ identity and ORF1 protein sequences share $\geq 15\%$ identity. Viruses identified in this study are noted with a red star. RdRP: RNA-dependent RNA Polymerase; ORF: Open Reading Frame; HP: Hypothetical Protein.

<https://doi.org/10.1371/journal.pone.0219207.g008>

these viruses together, but separate from other known (+)ssRNA viruses (Fig 9D). An additional threshold of 95% would distinguish each sequence evaluated here as a unique viral species; however, as the pairwise percent identity between AhoYV1 and AfSV2 of 94% is somewhat higher than is generally observed for unique species, a lower threshold may be more appropriate. Identification of additional sequences belonging to this group of viruses will further refine these sequence-based characteristics.

Another feature common to these viruses is the presence of at least one additional virus within the fungal host. Co-incident infection is not an uncommon occurrence in fungi; however, it has been demonstrated for two yadokariviruses, AfSV2 and YkV1, that isolated virus particles contain a second virus, related to Totiviruses [83,84]. Further, Zhang and colleagues demonstrated that YkV1 is *trans*-encapsidated with the capsid protein from the other virus, YnV1 [84]. As caliciviruses and totiviruses share a distant but strongly supported ancestry following diversification of the picorna-like superfamily [85], and both AfSV2 and YkV1 are present within virions encapsidated by totivirus capsid proteins, it is intriguing to speculate about the mechanisms at play, and the requirements for such an interaction. A totivirus was also identified in the *A. homomorphus* sample (AhoTV1, discussed above), however only a partivirus (PdPV1-HSF6) was identified along with PdYV1. For the remaining two members of this group of viruses, it is not possible to know the identity of the putative coat-protein donor, as PaFIV1 was found to co-infect *P. aurantiogriseum* along with a fusarivirus, totivirus, unclassified virus, and partivirus [4] and FpMyV2 was identified in a fungal strain that contained 13 other viruses from a variety of families [52].

Further study of the fungi that host a member of this group of viruses will answer questions regarding the relationship between the capsid-less virus and the capsid-donor, and may provide an interesting system for studying the evolution of a capsid-less virus. Due to the distinct phylogenetic relationship between these six identified and known (+)ssRNA viruses, along with the unique lifestyle, we support the recommendations to create a new virus family, named *Yadokariviridae* after YkV1, where *yadokari*, is the Japanese word for hermit crab [82,86].

***Narnaviridae* and *Ourmia*-like virus.** Members of the family *Narnaviridae* have single molecule, positive-strand RNA genomes ranging in length from 2.3 to 2.9 kb and encode an RdRP as the only polypeptide [87]. There are currently two genera within this family: *Narnavirus* and *Mitovirus*, where mitoviruses are characterized as viruses that replicate within the mitochondria of filamentous fungal hosts [87]. Indeed, a non-interrupted ORF is only predicted in these viruses when using the mitochondrial codon usage where tryptophan is encoded by UGA [88]. Here we describe five new viruses with similarity to members of the genus *Mitovirus* from RNA-seq datasets from four fungal hosts: *Setosphaeria turcica mitovirus 1* (StMV1) (PRJNA250530; unpublished), *Colletotrichum falcatum mitovirus 1* (CfMV1) (PRJNA272832; [89]), *Loramycetes juncicola mitovirus 1* (LjMV1) (PRJNA372853;

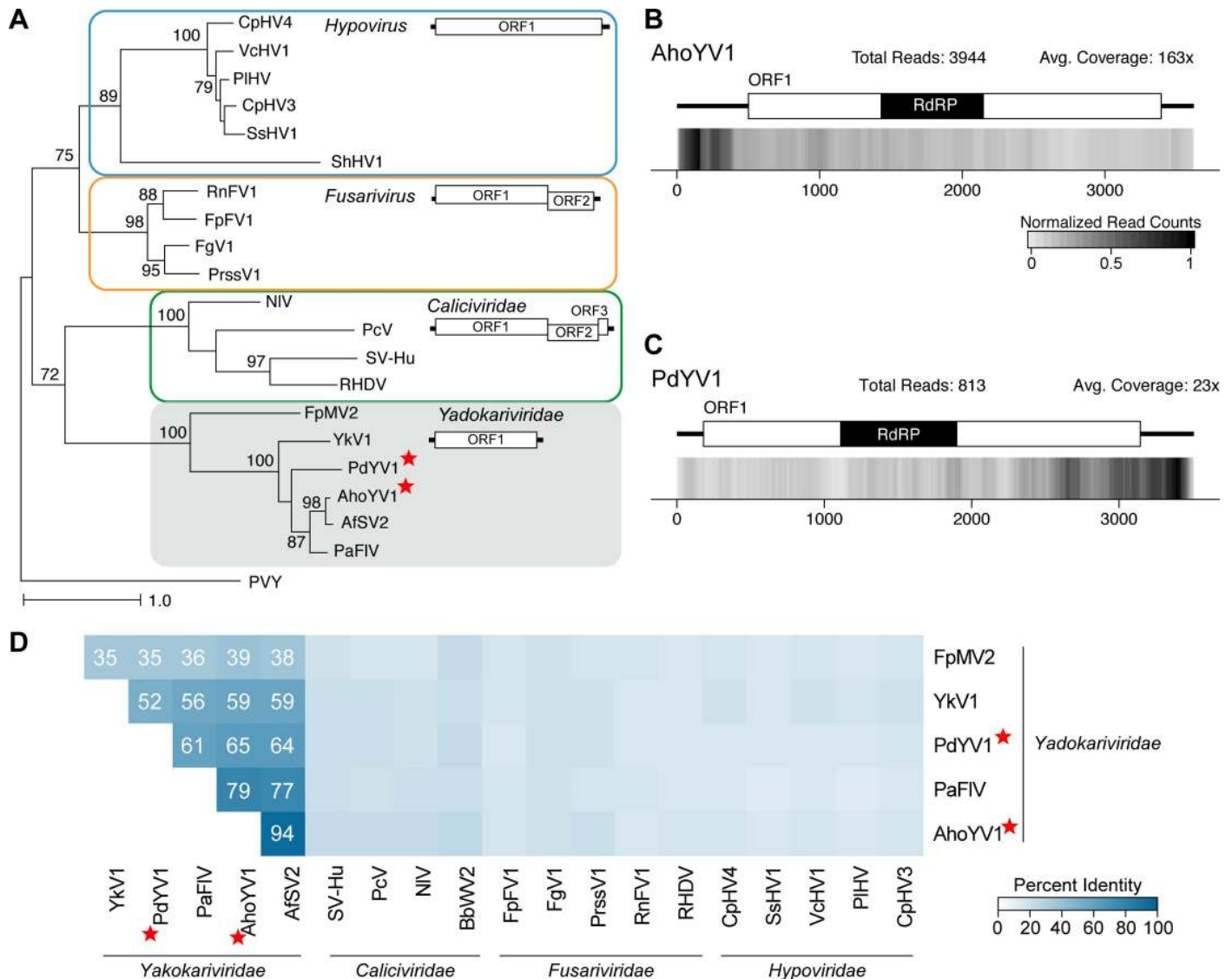


Fig 9. Analysis of *Aspergillus homomorphus yadokarivirus 1* (AhoYV1) and *Penicillium digitatum yadokarivirus 1* (PdYV1) viral genome sequences, members of the proposed family *Yadokariviridae*. (A) A phylogram of the RdRP coding sequence from members of the *Yadokariviridae* family and select (+)ssRNA viruses from the families *Caliciviridae*, *Fusariviridae*, and *Hypoviridae*; sequence from *Potato Virus Y* (PVY) serves as the outgroup. Virus names have been abbreviated; full names and accession numbers for protein sequences used in the alignment are in S2 Table. Phylogenetic tree was generated with RAxML, using the LG+G+I amino acid substitution model. Scale bar represents 1.0 amino acid substitutions per site and numbers at the nodes indicate bootstrap support over 70% (1000 replicates). To the right of the phylogram are representative diagrams of the genome structures for each family. (B and C) Density plot of read counts per nucleotide across the viral genome for (B) AhoYV1 and (C) PdYV1. Graphical depiction of the genome structure, the predicted single ORF and the RdRP motif are along the top and a heatmap of normalized read counts is on the bottom. Read counts are normalized to a 0–1 scale, where the maximum read count for a given genome equals 1. Legend for (C) is the same as the legend found in (B). Total reads mapped to the virus genome and average read coverage are noted for each virus. (D) Percent identity matrix generated by Clustal-Omega 2.1 using the RdRP motif sequences from the viruses analyzed in (A). For the sake of clarity, only the top half of the matrix is included and the 100% values along the diagonal have been removed. Percent identities greater than 25% are indicated. Viruses identified in this study are noted with a red star. RdRP: RNA-dependent RNA Polymerase; ORF: Open Reading Frame.

<https://doi.org/10.1371/journal.pone.0219207.g009>

unpublished), and *Ophiocordyceps sinensis mitovirus 1* and 2 (OsMV1, OsMV2) (PRJNA 292632; [21]).

The five mitoviruses have genome sizes within the expected range (Table 1), and a single ORF is present only when using the mitochondrial codon table. The percent of A+U in the coding strand for these five viruses ranges from 53% to 70%; characteristically, mitoviruses

have 62 to 73% A+U [87], although the percent observed in recently identified viruses is as low as 53% [5]. The predicted protein within each genome contains a Mitovir_RNA_pol domain, with a length ranging from 652 to 716 aa. Alignment of the RNA-seq reads to the viral genomes demonstrates end to end coverage for each virus, with average coverage per nucleotide ranging from 51x for OsMV1 to 540x for LjMV1 (S8D Fig).

A phylogenetic analysis of the RdRP coding sequence demonstrates strong support for all five newly identified viruses among members of the genus *Mitovirus* (Fig 7B). Criteria for species demarcation within the genus *Mitovirus* is not yet fully defined, however, viruses with greater than 90% identity within the RdRP protein sequences are considered the same species [42]. Here we observe that StMV1, CfMV1, OsMV1, OsMV2, and LjMV1 have less than 52% identity to any currently available mitovirus sequence (S8B Fig). Taken together, the above characteristics demonstrate that these five viruses are new species within the genus *Mitovirus*.

Ourmiaviruses are also positive-strand RNA viruses, composed of three genome segments that encode an RdRP, movement protein, and coat protein, isolated from plant hosts [41,90]. Despite usage of standard eukaryotic codons, phylogenetically the RdRP protein sequence is most closely related to that of *Narnavirus* [41]. Recently viruses have been identified from fungal hosts with sequence similarity to that of the plant-hosted ourmiaviruses. One such virus was identified in this study: *Aspergillus neoniger ourmia-like virus 1* (AnOIV1) (PRJNA250996; unpublished). The single segment, 2.7 kb genome encodes a single polyprotein that contains an RdRP motif (Table 1). Alignment of the RNA-seq data to the genome revealed an average coverage of 1007x (S8C Fig).

A phylogenetic analysis of this virus, along with related ourmia-like viruses from other fungi, plant-hosted ourmiaviruses, and mitoviruses illustrates a well-supported clade containing AnOIV1 along with *Phomopsis longicolla RNA virus 1* (PIRV1) [91] and *Botrytis ourmia-like virus HAZ2-3* (BoV) [92] (Fig 7B). A percent identity analysis of the fungal and plant viruses found in Fig 7B demonstrates a percent identity of approximately 50% among the recognized *Ourmiaviruses*, while the fungal-hosted viruses shared from 17 to 37% identity (S8A Fig), with AnOIV1 and PIRV1 the most similar. As there is limited sequence similarity, and AnOIV1 was identified from a new fungal host, it would appear that this is a new species within this yet-to-be fully characterized viral genus. Identification of additional *Ourmiaviruses* and related mycoviruses will shed further light on the specific relationship between these plant and fungal viruses.

Hypoviridae. The *Hypoviridae* family consists of positive sense, single-stranded RNA viruses that are approximately 9.1 to 12.7 kb in length, containing either one long ORF or two ORFs [93]. Further, these viruses are capsidless and use a host-derived, trans-Golgi network for genome replication [94,95]. Infection by Hypoviruses can often result in reduced fungal virulence, a hallmark symptom of early members of the *Hypoviridae* family.

Hypoviruses identified from four transcriptomic datasets are described here: *Sclerotinia homoeocarpa hypovirus 1* (ShHV1) (PRJNA167556; [25]), *Trichoderma asperellum hypovirus 1* (TaHV1) (PRJNA261111; [27]), *Setosphaeria turcica hypovirus 1* (StHV1) (PRJNA250530; unpublished), and *Fusarium graminearum hypovirus 1* (FgHV1) (PRJNA263651; [14]). The virus FgHV1 was expected, as the sequencing experiment featured *F. graminearum* strain HN10 infected with FgHV1 [14]; to our knowledge, this is the first mycovirus identified from *S. turcica* and *T. asperellum*, and the first hypovirus from *S. homoeocarpa*.

There are currently four species officially recognized in the *Hypoviridae* family: *Cryphonectria hypovirus 1, 2, 3, and 4* (CHV1, CHV2, CHV3, and CHV4), although several other related viruses have recently been identified from eight other fungal species. Of the currently recognized hypoviruses two clades form that appear to be distantly related, whereby CHV1 and CHV2 are most closely related and CHV3 and CHV4 are together in a separate grouping. To

this end, two new genera have been proposed, *Alphahypovirus* [96] and *Betahypovirus* [97], where CHV1 and CHV2 are classified in the former and CHV3 and CHV4 in the later [93]. A phylogenetic analysis using the RdRP protein sequence of the four viruses identified here, along with known and related hypoviruses, revealed that FgHV1, ShHFV1, and TaHV1 groups with the *Alphahypoviruses* while StHV1 was found with the *Betahypoviruses* (Fig 7C). A further analysis of these viral sequences, described below within the context of the two recently proposed genera, determined that ShHV1, StHV1, and TaHV1 are new species within the *Hypovirus* genus.

Alphahypovirus. Initially, members of *Alphahypovirus* were characterized as having two ORFs, where the second ORF contained protease, RdRP, and Helicase domains. This included CHV1 [98], CHV2 [99], FgHV1 [96], and *Rosellinia necatrix hypovirus 2* (RnHV2) [100]. However, identification of additional viruses has revealed members with a single, large ORF, including *Fusarium langsethiae hypovirus 1* (FlHV1) [101], *Fusarium graminearum hypovirus 2* (FgHV2) [96], *Fusarium poae hypovirus 1* (FpHV1) [52], and *Rosellinia necatrix hypovirus 1* (RnHV1) [100]. Independent of the number of ORFs, alphahypoviruses have large genomes, with lengths ranging from 12.5 to 14.4 kb. Similarly, newly identified viruses ShHV1 and TaHV1 have genome lengths of 12.4 and 14.2 kb respectively, and single-ORF genome structures (S9A Fig, Table 1).

For a new species to be classified, criteria generally include host species and percent identity with other identified species. Official criteria do not yet exist for the family *Hypoviridae*, and separate criteria for the two proposed genera may be appropriate given the lack of sequence homology between the two groups. Previously it has been observed that CHV1 and CHV2, while both isolated from *Cryphonectria parasitica*, demonstrate only ~67% percent identity within the RdRP amino acid sequence. Expanding this analysis to all viruses within the genus *Alphahypovirus* reveals a maximal identity of 90%, between FgHV2 and FpHV1 (S9B Fig); this excludes the 99% identity observed between the two sequences of FgHV1 which are known to be the same species of virus [14]. Interestingly, there is a cluster of similar hypoviruses from three different *Fusarium* species, FlHV1, FpHV1, and FgHV2, which excludes the fourth *Fusarium* hypovirus, FgHV1. An appropriate threshold for evaluating new species of *Alphahypovirus* remains to be determined; however as the maximum shared identity with existing references sequences is less than 51% for ShHV1 and 31% for TaHV1, and both viruses were identified from unique fungal hosts, these appear to be new species within the family *Hypovirus*.

A final analysis of the three alphahypoviruses identified from this study was the alignment of RNA-seq reads to the viral genomes. Total reads that aligned range from 7,913 for ShHV1 to 323,433 for FgHV1, and in all three instances, the peak coverage is closest to the 3' end of the genome (S9A Fig).

Betahypovirus. Viruses in the *Betahypovirus* genus generally encode one ORF that contains protease, RdRp, Helicase, and UGT domains [97] and tend to have shorter genomes than the *Alphahypoviruses*, at approximately 9 to 11 kb in length. These same characteristics are also observed in StHV1, which has a length of 9,163 bp and a polyprotein predicted to encode a papain-like cysteine protease domain along with a UGT, peptidase-C97, RdRP, and Helicase-HA2 domain (S9C Fig, Table 1). Additionally, the 5' end of the 5'UTR sequence has high sequence identity across viral genomes from this genus; a multiple sequence alignment of this region from StHV1 and related *Betahypoviruses* reveals similarly high conservation within the StHV1 genome, including a 15-nt sequence that is present in all viruses analyzed (CTGGTTTACTCTG) (S9D Fig). Characterization of the StHV1 sequence also included alignment of the RNA-seq reads to the viral genome, resulting in a total of 376,89 reads aligned, and an average coverage per nucleotide of 617x (S9C Fig).

A percent identity analysis of the RdRP amino acid sequences of viruses within the genus *Betahypovirus* reveals a maximal identity of 68%, between PIHV and CHV3 (S9B Fig) while StHV1 shares 46% to 54% identity with other viruses. As such, a threshold of 70% identity within the RdRP amino acid is suggested as a criteria to distinguish between different species of *Betahypoviruses*.

Fusariviridae. Members of the proposed family *Fusariviridae* have mono-segmented, plus strand, single-stranded RNA genomes that encode two to four ORFs. ORF1 is the largest and contains the RdRP and Helicase motifs, while the subsequent and second largest ORF encodes an unknown protein, often containing an SMC motif. One to two additional small proteins have been predicted at the 3' end of some virus genomes. Here we have identified seven new viruses that group with fusariviruses.

Aspergillus ellipticus fusarivirus 1 (AeFV1) (PRJNA250911; unpublished), *Morchella importuna fusarivirus 1* (MiFV1) (PRJNA372858; unpublished), *Neurospora discreta fusarivirus 1* (NdFV1) (PRJNA257829; [20]), and *Rutstroemia firma fusarivirus 1* (RfFV1) (PRJNA372878; unpublished) are to our knowledge, the first viruses to be identified in these species. *Sclerotinia homoeocarpa fusarivirus 1* (ShFV1) (PRJNA167556; [25]) and *Gaeumannomyces tritici fusarivirus 1* (GtFV1) (PRJNA268052; [16]) are the first fusariviruses identified in these fungi, and this is the first full-length description of *Zymoseptoria tritici fusarivirus 1* (ZtFV1) (PRJNA237967; [30–32] PRJEB8798; [99]) (PRJNA253135; [[30–32]]) (PRJNA179083; unpublished).

The genome of MiFV1 encodes three predicted ORFs, where the first contains both the RdRP and helicase motifs, the second contains no predicted domains, and the third ORF contains an SMC (structural maintenance of chromosomes) domain (S10C Fig). The genomic structure of the other six viruses identified in this study encode two predicted ORFs, again with the first encoding RdRP and helicase motifs (S10C Fig). ORF2 of ZtFV1 contains a predicted Spc7 domain (found in kinetochore proteins), a domain also found in *Rosellinia necatrix fusarivirus 1* (RnFV1) [102]. Within ORF2 of AeFV1, GtFV1, NdFV1 and RfFV1 is a predicted SMC domain, while no functional domain is predicted within ORF2 of ShFV1. A previous phylogenetic analysis determined that the closest relatives to the viral SMCs are the proteins from the eukaryotic SMC5 and SMC6 subfamily and that because the SMC domain is widely present throughout members of the *Fusariviridae* family, a common ancestral origin is probable [103].

A phylogenetic analysis of the RdRP amino acid sequence of these seven viruses and twelve other published fusariviruses reveals two strongly supported clusters (S10A Fig). “Group 1” contains 14 viruses while “Group 2” contains the remaining five fusariviruses. The genome structures of the two clusters suggests some general patterns; all members of Group 1 have a predicted functional domain within the second largest ORF, either an SMC or an Spc7 domain as described above. Although, FgV1 encodes three ORFs, unlike related viruses that encode just two. Conversely, viruses within Group 2 generally do not have a functional domain predicted within ORF2, the exception is MiFV1 which contains an SMC domain. Further, three ORFs are predicted for four of the five members, with ShFV1 being the exception. Identification of additional viral genomes that belong to this family will help to elucidate the support for these separate clusters and may reveal additional distinguishing characteristics.

As *Fusariviridae* is not yet an officially recognized family of viruses, specific characteristics for distinguishing new viral species or genera have yet to be established. In addition to phylogenetic analyses and identity of fungal host species, percent identity across the protein coding sequence of the RdRP-containing ORF is frequently used as a classifier. To date, fusariviruses previously characterized have all been isolated from unique fungal species, including two different species of *Fusarium*, and none share greater than 57% identity with one another (S10B

Fig). Viruses identified in this study are also from unique fungal species, including ShFV1 which was isolated from a different species than *Sclerotinia sclerotiorum fusarivirus 1* (SsFV1) [104]. Using the data available to date, two criteria are recommended to qualify viruses as new species: a unique fungal host and a percent identity of the RdRP sequence $\leq 60\%$ in comparison to existing fungal sequences.

The ZtFV1 sequence described here was identified within the SRA sample PRJNA179083, however there are three other RNA-seq experiments available for this organism from three other research groups. All four sequencing samples contain the same fusarivirus; indeed a full length sequence was isolated from samples from both the Max Planck Institute for Terrestrial Microbiology (PRJNA237967; [31]) and Rothamsted Research (PRJEB8798; [30]). The sample from the Institut national de la recherche agronomique (INRA) (PRJNA253135; [32]) yielded a contig covering 76% of the full length, while a bowtie2 alignment of reads from this sample to the full-length sequence demonstrated reads across the full length sequence. Indeed, the average read coverage per nucleotide using reads from each of the four samples was high, ranging from 3548x to 15278x (S10C Fig). As the length of 5' poly(G) and 3' poly(A) sequences differed slightly among the four samples, the most parsimonious version was selected and deposited into GenBank (Table 1). Other than these differences, no SNPs were observed among the four samples. While the only information related to when these samples were sequenced is the date the data was made public the range in dates, between November 2012 and June 2015, suggests that the viral genome sequence may have remained unchanged across multiple years. An interesting follow-up experiment would be a re-sequencing of one of these strains by the same lab to observe the viral genome sequence following multiple years of culturing within the fungal host.

MiRV1. In addition to the three endornaviruses and one fusarivirus described above, a fifth viral sequence was identified in the *M. importuna* sample (PRJNA372858; unpublished). The single-segment sequence was determined to be 10,132 nt in length, with a single predicted ORF of 3,295 aa (Table 1). A conserved domain search revealed a viral methyltransferase domain and a viral helicase domain in addition to the RdRP_2 motif similar to positive-strand RNA viruses (S11C Fig). A BLAST-P search with the predicted protein sequence returned a single full length mycovirus sequence, *Sclerotinia sclerotiorum RNA virus L* (SsRV-L) [105], in addition to matches to Hepatitis Virus E and Cordoba virus. Other top hits include partial sequences corresponding to the RdRP domain of *Rhizoctonia solani RNA virus 1, 2, and 3* (RsRV-1, RsRV-2, RsRV-3) [106]. Both MiRV1 and SsRV-L contain a single ORF with methyltransferase, helicase and RdRP domains, and neither contain peptidase motifs [105].

Previous phylogenetic analyses of SsRV-L and the three viruses from *R. solani* demonstrated a relationship between these viruses and members of the alpha-like virus superfamily, specifically viruses from the *Hepeviridae* and *Alphatetraviridae* families [105,106]. *Endornaviridae*, whose members infect both fungi and plants, are also a family of alpha-like viruses [107], but appear to be more distantly related to this emerging group of mycoviruses. Expanding the phylogenetic analysis to include MiRV1 confirms that MiRV1 is a member of this new clade of mycoviruses (S11A Fig). Further, a multiple sequence alignment of the RdRP motif sequences demonstrated the presence of the eight conserved RdRP motifs common among (+)ssRNA viruses, along with certain residues within these motifs were shared only among the mycoviruses that group with MiRV1 (S11B Fig). The low percent identity between MiRV1 and the related mycoviruses indicates that this is a new species of virus (S11D Fig). Identification of additional related viral sequences will help elucidate the relationship of MiRV1 and the rest of the Alpha-like viruses.

Unknown comovirus-like sequences. An RdRP-containing contig with a top match to Turnip Ringspot Virus (TRV) was identified in sequencing samples from *Colletotrichum*

tofieldiae (PRJNA287627; [108]) and *Zymoseptoria brevis* (PRJNA277175; [109]). Further inquiry also identified the second genomic segment of TRV, which encodes the movement and coat proteins, from both the *C. tofieldiae* and *Z. brevis* Trinity contigs. The coding sequences of RNA1 and RNA2 are 98.5% and 99.7% identical between the two isolates suggesting a close relationship between the sequences, despite the unrelated hosts.

As Turnip Ringspot Virus (genus: *Comovirus*; family: *Secoviridae*) can infect *Brassicaceae* [110], and a fungal host has not yet been described for viruses from this family, we postulate that within the *C. tofieldiae* sample, this sequence could be due to cross contamination from *Arabidopsis*. This BioProject included both media-grown fungi and an infection series of *Arabidopsis thaliana*; the specific SRA record used in this analysis was from a media-grown sample. Further analysis of the infection series determined that the viral sequence is identifiable in 19 of 24 (79%) of the mock-infected *A. thaliana* samples, 19 of 24 (79%) of the *A. thaliana* plus *C. tofieldiae* samples, and two of three fungus-only samples. An explanation for the uneven distribution of the viral sequences among these samples remains to be determined. Conversely, the *Z. brevis* sample was part of a media-grown, fungus-only transcriptomic experiment and as such, a putative plant-based source of the TRV-like sequences is not readily apparent. Interestingly, while the *Z. brevis* sample is from Christian-Albrechts University of Kiel (Kiel, Germany) and the *C. tofieldiae* sample is from the Max Planck Institute for Plant Breeding Research (Kohn, Germany), both were sequenced at the Max Planck Genome Center (Cologne, Germany) near the same time and on the same machine based on the identifiers in the sequencing headers. However, cross contamination between the samples is unlikely as it would be expected the observed viral sequences would be 100% identical which is not the case. At this junction, it is not possible to confirm the host or source of the viral genome sequences identified in the *C. tofieldiae* and *Z. brevis* samples, nor determine if these represent a new group of mycoviruses that are related to Comoviruses.

Conclusions

Here we describe the power of a systematic approach for analyzing public RNA-seq datasets to identify RNA virus genomes. Application of this bioinformatics pipeline to 569 transcriptomic datasets from 312 species of Pezizomycotina fungi identified 59 full-length viral genomes, 52 of which were determined to be new viruses.

The success of the pipeline was due to (a) the ability to fully assemble a viral RdRP coding sequence from as few as 136 reads (or 12x coverage) and (b) the strong signature of the viral RdRP domain. Benefits of this analysis include suitability to either rRNA-depleted total RNA or poly-A purified RNA. As viral particles enrichment is not required, which would select against non-particle forming viruses, using RNA-seq data to identify viral sequences can easily be applied to commonplace transcriptomic experiments. Notable limitations of this approach include bias for viruses with a recognizable RdRP domain, which automatically excludes DNA viruses and may select against (-)ssRNA viruses. Further, robust analyses are required following the identification of putative viral sequences to confirm homology and relatedness to the set of known viruses.

This analysis is not just an academic exercise in computational pipeline development and execution. It also works to answer the biological question, what fungi are infected by what viruses? It is becoming increasingly clear that there is a dynamic range of responses due to viral infection, outside of the well documented asymptomatic or hypovirulence phenotypes, that reflect the complex relationships present during multipartite interactions. A recent review by Roossinck [111] highlights a number of mutualistic or symbiogenic relationships between viruses and their hosts. This includes killer yeasts: yeast that are infected with a virus which

produces toxins that kill competitors to the host fungi while conferring resistance to the host [112]. Another example is the tripartite interaction between the tropical panic grass *Dichanthelium lanuginosum*, the endophytic fungi *Curvularia protuberata* and *Curvularia* thermal tolerance virus, the virus that infects *C. protuberata*. Thermal tolerance to high soil temperatures is only conferred to the plant when the fungi is present and is infected with the mycovirus [113]. Another phenotype, different from canonical hypovirulence, is the reduction in fungicide resistance observed following viral infection of prochloraz-resistant strain of *Penicillium digitatum* [114].

Multipartite relationships also include virus-virus interactions within a host as many fungi are concurrently infected by more than one mycovirus; here we identified over 10 such fungi. Recently, it was described that a positive effect on accumulation of a viral genome was observed when a second, unrelated virus was present in *Rosellinia necatrix* [84]. Conversely, coinfection of *C. parasitica* by a totivirus and a suppressor-deficient hypovirus lead to induction of antiviral defense genes and ultimately inhibition of totivirus replication [115]. Thus evaluating the impact on fungal phenotype during concurrent infections is likely host-virus-virus specific.

Asymptomatic or latent infections can downplay the active antiviral processes at work within the fungus and the important contributions a successful antiviral defense has on fungal phenotypes. Small RNA-mediated antiviral defense systems, crucial for preventing a deleterious effect on fungal health, have been identified in *Cryphonectria parasitica* [116], *Colletotrichum higginsianum* [8], and *Sclerotinia sclerotiorum* [117]. Further, it can be assumed that these antiviral defenses have contributed to the under-identification of mycoviral infections in fungi, particularly as the RNA silencing pathways in most fungal hosts identified in this study have yet to be evaluated.

This systematic analysis also highlights the uneven distribution of sequencing data among members of the Pezizomycotina subphylum. While strides are being made, particularly due to large-scale approaches such as the 1000 Fungal Genomes Project, a fuller understanding of the mycoviral landscape and therefore the effect a virus has on fungal phenotype requires continuous inquiry. As the number of Pezizomycotina fungi is estimated to be greater than 32,000 [118] and currently less than 500 fungal species have been queried, it stands to reason there remains a wide range of viral diversity yet to be identified and characterized. The challenges ahead with respect to the current repository of sequencing data is that this study focused solely on the subphylum Pezizomycotina, meaning other members within the Dikarya and other subphyla remain unexamined. And that beyond the fungal kingdom are the many more RNA-seq datasets from plants and animals. The computational requirements to investigate these tens of thousands of RNA-seq datasets are not small and identifying suitable samples at the start of the pipeline as well as the bioinformatic characterization of candidate contigs at the end of the pipeline currently requires a hands-on approach. There are also opportunities available for expanding the scope of this analysis, specifically the inclusion of single-molecule and/or long-read sequencing data that includes reads >1 kb in length; these may provide an even more direct route to virus detection as an entire viral genome could be sequenced as a single read. Finally, looking beyond the data that already exists, we highly recommend that an analysis of sequencing data for the presence of viral RdRP signatures becomes part of all routine bioinformatics pipelines.

Materials and methods

Identifying publicly available fungal RNA-seq datasets

Programs available via the Entrez Programming Utilities (“E-utilities”) were used to search the NCBI Short Read Archive (SRA) for datasets that matched the search criteria. For ease of

searching, the fungal kingdom was divided into three phylum level searches: ascomycota (sub-phylum Pezizomycotina), basidiomycota, and other. General search criteria used for all phylum-level searches included the phrase “NOT LS454 NOT PACBIO NOT ABI_SOLID AND biomol rna [PROP]” to identify RNA-seq datasets from Illumina (or Solexa) instruments. The E-utilities commands `esearch` and `efetch` combined to create a comma-separated table of all matching SRA records.

Pipeline for identification of viral-like sequences

The output file from the E-utilities search command was manually parsed to create a unique list of all BioProjects and fungal species. Multiple fungal species found under one BioProject number were split into separate entries. Most BioProjects included more than one sample; for example different treatments or wild-type versus mutant. In general, the SRA sample selected for analysis was a wild-type-like strain grown on media. When available, the fungal genome sequence was either downloaded from NCBI or Joint Genome Institute (JGI).

The computational steps for identifying viral sequences started with a downloaded raw reads file from SRA (Fig 10). The SRA Toolkit program `fastq-dump` (v2.5.7) was used to download and convert the SRA file to fastQ format. When a fungal genome sequence was available, RNA-seq reads were aligned to the genome using `bowtie2` (version 2.2.9) [119], and the unaligned reads were captured in a separate file using the flag ‘-un’. A *de novo* assembly of these unaligned reads was created using `Trinity` (v2.1.1) [120] and resultant contigs were analyzed for open reading frames (ORFs) by `TransDecoder` (version 2.1) [121]. Finally, the predicted protein sequences were queried against a custom database of viral RNA dependent RNA polymerases (RdRPs) using `HMMscan` (version 3.1b2; [122]). In general, when the data was paired end, the data present within pair one was sufficient data for identification of viral contigs, however for the sake of completeness, following identification of a viral genome, the pipeline was re-run using both pairs. Additionally, a step was added between `fastq-dump` and `bowtie2` whereby the reads containing Illumina adaptor sequences were removed using `cutadapt` (v1.18) [123]. When using both paired-end data files, due to the nature of the ‘-un’ flag in `bowtie2`, the pairs were treated as individual single-end files for the creation of the unaligned reads file used as input for `Trinity`.

Assembly of the full-length viral sequence from *Trichoderma harzianum* (PRJNA216008), was particularly recalcitrant and as such, all three paired-end datasets were used in the pipeline for identification of ThAV1.

Contig extension and identification

Contigs identified containing putative viral RdRP domains were further analyzed by first aligning the sequence to the ‘nr’ database at NCBI using `BLASTX`. Some viruses were determined to have multipartite genomes, for example members of the *Partitiviridae* or *Chrysoviridae* families. To identify the other genomic segments, first a `BLASTP` database of the translated `Trinity` contigs was created and then queried with the protein coding sequence from the additional genomic segment from the top `BLASTX` match from the RdRP query of the ‘nr’ database. S2 Table summarizes the viral RdRP sequences identified and, where applicable, which known sequences were used to query the `Trinity` assembly for additional contigs.

For some other viral sequences, the RdRP-containing segment identified was possibly truncated since the predicted ORF(s) extended uninterrupted to either the 5’- or 3’-most edge, indicating either the start or stop codon may be missing. In instances when the in-hand sequence was near the expected length, based off of the length of the top `BLASTX` matches, reads that overlapped the truncated end were identified and `MAFFT` alignment of these reads

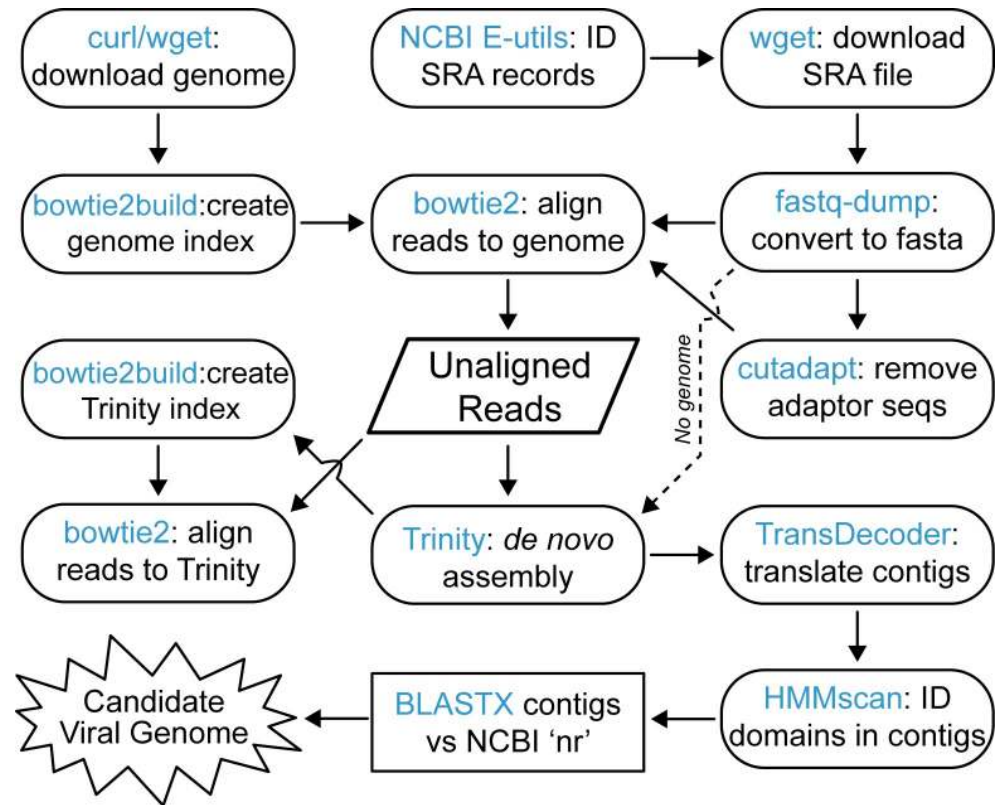


Fig 10. Bioinformatics steps of the virus-like RdRP discovery pipeline. Computational program names are highlighted in blue. Reads used in the bowtie2 alignment step were either as-is or adaptor-trimmed by cutadapt; details in the Materials and Methods. The dashed line indicates the workflow for fungi where no genome sequence was available; instead of running bowtie2, the reads were fed directly into a Trinity *de novo* assembly.

<https://doi.org/10.1371/journal.pone.0219207.g010>

to the RdRP-containing contig allowed for the extension of the contig to its final length. In other instances, contigs containing only the RdRP-containing ORF that were predicted to originate from mono-segmented viruses, whereby a second ORF is present within the genome, were identified, for example *Totiviruses*. For these sequences, the protein coding sequence of the missing ORF was identified from the top BLASTX match and a custom BLASTP database of the Trinity protein coding sequences was queried. After identifying the contig containing the second ORF, reads were identified to stitch together the two contigs to create the final viral genomic sequence. Once the final viral genomic sequence was identified, ORFs were predicted using ORFinder at NCBI [124].

Bioinformatic tools and analyses

Sankey diagram. The online tool SankeyMATIC [125] was used generate the Sankey diagram in Fig 1. Data tables used as input are available through the FigShare FileSet (DOI: [10.6084/m9.figshare.7476359](https://doi.org/10.6084/m9.figshare.7476359)).

Pseudoknot structures. The online tool DotKnot was used to predict pseudoknots in certain viral genome sequences [126,127]. PseudoViewer 3.0 was used for visualization of the predicted pseudoknots [128].

Read distribution plots. Distribution of reads along the viral genome segment(s) was visualized by first aligning the reads to the viral genome using bowtie2, then calculating the hits per nucleotide along the viral sequence(s) using the custom Perl script

plotRegion_RNAseq_hitsPerNT.pl. The resulting output table was the input for an R script to create hits per nucleotide heat maps for each viral segment. The R code for each plot, along with the input data table, is available through the FigShare FileSet (DOI: [10.6084/m9.figshare.7476359](https://doi.org/10.6084/m9.figshare.7476359)).

Phylogenetic analysis. To identify related sequences for phylogenetic analysis, the predicted amino acid sequence for the RdRP was used to query the ‘nr’ database at NCBI. Protein sequences of the top hits, generally with e-values approaching or equal to 0.0, were downloaded. Partial sequences were excluded. Known viral sequences that appeared as the top hit for multiple viruses identified in this study were deduplicated. The complete list of Accession IDs used for phylogenetic analysis is available in [S2 Table](#).

To build the trees, amino acid sequences were aligned with MAFFT [129], and resultant alignments were degapped and trimmed using the tool trimAL [130] with a gap threshold of 0.9 which allowed for a gap in up to 10% of the sequences at a given position. For seven of the nine phylogenetic trees, the number of alignment sites that remained after trimming ranged between 300 and 400 amino acids; a slightly smaller range was observed for the MiRV1 tree (242 to 252 amino acids), and a larger range of alignment sites remained for the *Hypoviridae* tree (1292 to 1592 amino acids). The unaligned, untrimmed fasta files and the aligned, trimmed fasta files are available through the FigShare FileSet (DOI: [10.6084/m9.figshare.7476359](https://doi.org/10.6084/m9.figshare.7476359)). The later files were used as input for the creation of maximum likelihood phylogenetic trees using RAxML version 8.2.4 [131] and the LG+G+I amino acid substitution model. RAxML tree files with branch lengths were viewed in Dendroscope version 3.5.7 [132] and exported as PDF for figure preparation; both the RAxML and the PDF files are available within the FigShare FileSet (DOI: [10.6084/m9.figshare.7476359](https://doi.org/10.6084/m9.figshare.7476359)).

Percent identity matrices. Percent identities among protein coding sequences, either full length or motif-specific, were determined via multiple sequence alignment using Clustal Omega [133]. In instances where two different proteins were evaluated and then combined into one matrix, such as the RdRP protein and the Coat Protein, first the RdRP protein sequences were analyzed; the order of the output in this percent identity matrix was then applied to the second set of protein sequences to be analyzed and the option “Order: Input” was selected in Clustal-Omega. Percent identity matrices were converted to heat map plots using a custom R script. This script along with the input matrices are available through the FigShare FileSet (DOI: [10.6084/m9.figshare.7476359](https://doi.org/10.6084/m9.figshare.7476359)).

Supporting information

S1 Fig. Analyses of *Hortaea werneckii totivirus 1* (HwTV1), a virus from the *Totivirus* genus within the *Totiviridae* family. (A) Percent identity matrix, generated by Clustal-Omega 2.1, to select totiviruses. ICTV-recognized members of the genus are in bold. The top half of the matrix, in the blue scale, is the percent identity of RdRP protein sequences and the bottom half of the matrix, in the green scale, is the percent identity of CP protein sequences. For sake of clarity, the 100% identity along the diagonal has been removed and outlined stars are used to note junction point for the row and column of the virus identified. Identities $\geq 50\%$ are noted. Full names and accession numbers for protein sequences are in [S2 Table](#). (B) Slippery site and two pseudoknot structures predicted using DotKnot and visualized with PseudoViewer3. The putative slippery site heptamer (GGAUUUU) is highlighted in red and the CP stop codon (UAG) is in highlighted in orange. (C) Density plot of read counts per nucleotide across the genome of HwTV1. Graphical depiction of the genome structure, encoded genes and predicted protein motifs are on top and a heatmap of normalized read counts is on the bottom. The predicted slippery site and pseudoknot region is indicated by the dashed box.

Read counts are normalized to a 0–1 scale, where the maximum read count for the genome equals 1. Total reads mapped to the virus genome and average read coverage are noted. Viruses identified in this study are noted with a red star. RdRP: RNA-dependent RNA Polymerase; CP: Coat Protein; mCP: major coat protein domain of L-A virus. (TIF)

S2 Fig. Analyses of identified viral genomes belonging to the *Victorivirus* genus within the *Totiviridae* family. (A) Percent identity matrix, generated by Clustal-Omega 2.1, comparing identified viral sequences to select members of the genus *Victorivirus*. The top half of the matrix, in the blue scale, is the percent identity of RdRP protein sequences and the bottom half of the matrix, in the green scale, is the percent identity of CP protein sequences. For sake of clarity, the 100% identity along the diagonal has been removed and outlined stars are used to note junction point for the row and column of the viruses identified. Where the identified viral sequences have significant similarity to known viral sequences is indicated, specifically RdRP or CP sequences with similarity $\geq 60\%$. Virus names in bold are ICTV-recognized species within the genus. Full names and accession numbers for protein sequences are in [S2 Table](#). (B) Percent of alanine, glycine, and proline (A, G, P) amino acids in the CP coding sequence. The binned values were calculated for twenty-five known victoriviruses and ten known totiviruses, shown as dark and light grey lines respectively. The nine putative victoriviruses identified in this study were individually analyzed and plotted. (C and D) Density plot of read counts per nucleotide across the viral genome for the nine viruses identified. Graphical depiction of the genome structure, predicted ORFs, and putative motifs are illustrated on top and a heatmap of normalized read counts is on the bottom. Reads are normalized to a 0–1 scale, where the maximum read count for a given genome equals 1. Total reads mapped to the virus genome and average read coverage are noted for each virus. The overlap between the CP stop codon (*italics*) and RdRP start codon (**bold**) is also indicated. Red stars indicate viruses identified in this study. RdRP: RNA-dependent RNA Polymerase; CP: Coat Protein; ORF: Open Reading Frame. (TIF)

S3 Fig. Analyses of viral genomes for members of the genus *Alphapartitivirus*. (A) Percent identity matrix generated by Clustal-Omega 2.1 of viruses identified in this study and select alphapartitiviruses. The top half of the matrix, in the blue scale, is the percent identity of RdRP protein sequences and the bottom half of the matrix, in the green scale, is the percent identity of CP protein sequences. For sake of clarity, the 100% identity along the diagonal has been removed and outlined stars are used to note the junction for the row and column of the viruses identified. Identities above the species demarcation are noted: $\geq 90\%$ for RdRP and $> 80\%$ for CP. Names in bold are ICTV-member species, while names in *italics* are related, unclassified viruses. Full names and accession numbers for protein sequences are in [S2 Table](#). (B) Density plot of read counts per nucleotide across the viral genome for the five alphapartitiviruses identified in this study. Graphical depiction of the genome structures of RNA1 and RNA2, the predicted ORFs, and RdRP motif are on illustrated above the heatmap of normalized read counts. Reads are normalized to a 0–1 scale, where the maximum read count for a given genome equals 1. Total reads mapped to the virus genome and average read coverage are noted for each virus. (C) Alignment of 5'UTRs of RNA1 and RNA2 genome segments for each virus. Residues not shown in alignment are in parentheses. Blue shading indicates regions of 100% identity. Red stars note the viruses identified in this study. RdRP: RNA-dependent RNA Polymerase; CP: Coat Protein. (TIF)

S4 Fig. Analyses of identified viral genomes belonging to the genus *Betapartitivirus*. (A) Percent identity matrix, generated by Clustal-Omega 2.1, between identified viruses and select betapartitiviruses. The top half of the matrix, in the blue scale, is the percent identity of RdRP protein sequences and the bottom half of the matrix, in the green scale, is the percent identity of CP protein sequences. For sake of clarity, the 100% identity along the diagonal has been removed and outlined stars are used to note junction point for the row and column of the viruses identified. Where the percent similarity is above the cutoffs for species demarcation is noted, specifically >90% for RdRP sequences and >80% for CP sequences. Bolded names are ICTV-recognized member species. Full names and accession numbers for protein sequences are in [S2 Table](#). (B) Density plot of read counts per nucleotide across the viral genome for the two *Betapartitiviruses*. Graphical depiction of the genome structure of the two RNA genome segments, the predicted ORFs, and the RdRP motif are along the top and a heatmap of normalized read counts is on the bottom. Read counts are normalized to a 0–1 scale, where the maximum read count for a given genome equals 1. Total reads mapped to the virus genome and average read coverage are noted for each virus. (C) Alignment of 5'UTR sequences from the RNA1 and RNA2 genome segments for both viruses. Blue boxes highlight conserved nucleotides and numbers in parentheses are residues not shown. Viruses identified in this study are noted with a red star. RdRP: RNA-dependent RNA Polymerase; CP: Coat Protein. (TIF)

S5 Fig. Analyses of identified viral genomes belonging to the genus *Gammapartitivirus*. (A) Percent identity matrix, generated by Clustal-Omega 2.1, between identified viruses and select gammapartitiviruses. The top half of the matrix, in the blue scale, is the percent identity of RdRP protein sequences and the bottom half of the matrix, in the green scale, is the percent identity of CP protein sequences. For sake of clarity, the 100% identity along the diagonal has been removed and outlined stars are used to note junction point for the row and column of the viruses identified. Where sequences have a value greater than the species demarcation cutoff is noted, specifically >90% for RdRP and >80% for CP sequences. Bolded names are ICTV-recognized member species while italicized names are related, but unclassified viruses. Full names and accession numbers for protein sequences are in [S2 Table](#). (B) Density plot of read counts per nucleotide across the viral genome for the five gammapartitiviruses. Graphical depiction of the genome structure of the RNA segments, predicted ORFs, and RdRP motif are on top and a heatmap of normalized read counts is on the bottom. Read counts are normalized to a 0–1 scale, where the maximum read count for a given genome equals 1. Total reads mapped to the virus genome and average read coverage are noted for each virus. (C) Alignment of 5'UTR sequences of RNA1 and RNA2 for the identified viral sequences. Conserved nucleotides are highlighted with a blue box and number of nucleotides not shown in the alignment are noted in parentheses. Viruses identified in this study are noted with a red star. RdRP: RNA-dependent RNA Polymerase; CP: Coat Protein. (TIF)

S6 Fig. Analyses of viral genomes belonging to the *Chrysoviridae* family. (A) Percent identity matrix, generated by Clustal-Omega 2.1, between the chrysovirus identified in this study versus select members of the genus *Chrysovirus*. The top half of the matrix, in the blue scale, is the percent identity of RdRP protein sequences and the bottom half of the matrix, in the green scale, is the percent identity of CP protein sequences. For sake of clarity, the 100% identity along the diagonal has been removed and outlined stars are used to note junction point for the row and column of the viruses identified. Values above 70% for RdRP sequences and 53% for CP sequences are indicated. ICTV-member and unclassified viruses are noted in bold and italics respectively. Full names and accession numbers for protein sequences are in [S2 Table](#). (B

and C) Alignment of the 5'UTR sequences for the four genomic segments of BbCV1 (B) and PrCV1 (C). Two highly conserved regions are found in the 5'UTR; the first is a 40–75 nt region, highlighted in green, while the second region is rich in CAA repeats (underlined). (C) Density plot of read counts per nucleotide across the viral genome segments. Graphical depiction of the genome structure and predicted ORFs and motifs are on top and on the bottom is a heatmap of read counts normalized to a 0–1 scale, where the maximum read count for a given genome equals 1. Total reads mapped to the virus genome and average read coverage are noted. Red stars indicate viruses identified in this study. RdRP: RNA-dependent RNA Polymerase; CP: Coat Protein; HP: Hypothetical Protein.

(TIF)

S7 Fig. Analyses of identified viral genomes belonging to the family *Endornaviridae*. (A and C) Percent identity matrix, generated by Clustal-Omega 2.1, of mycoviruses in the genus *Alphaendornavirus* (A) and *Betaendornavirus* (C). The top half of both matrices, in the blue scale, is the percent identity of the RdRP motif sequences; the bottom half of the matrix, in the green scale, is the percent identity of the Viral Helicase 1 motif (A) and Viral methyltransferase motif (C). For sake of clarity, the 100% identity along the diagonal has been removed and outlined stars are used to note junction point for the row and column of the viruses identified in this study. Bolded names are ICTV-recognized member species. Full names and accession numbers for protein sequences are in [S2 Table](#). (B and D) Density plot of read counts per nucleotide across the viral genome for GeEV1 and MiEV2 (B) and MiEV1 and MiEV3 (D). Graphical depiction of the genome structure, predicted ORF, and encoded motifs is on the top and a heatmap of normalized read counts is on the bottom. Read counts are normalized to a 0–1 scale, where the maximum read count for a given genome equals 1. Total reads mapped to the virus genome and average read coverage are noted. Viruses identified in this study are noted with a red star. RdRP: RNA-dependent RNA Polymerase; VH: Viral Helicase 1; Vmt/Vm: Viral methyltransferase; DH: DEXDx Helicase.

(TIF)

S8 Fig. Analyses of identified viruses related to *Ourmiavirus* and *Narnaviridae*. (A and B) Percent identity matrix of the RdRP protein coding sequence, generated by Clustal-Omega 2.1, of ourmia-like mycoviruses (A) and the genus Mitovirus from the family *Narnaviridae* (B). For the sake of clarity, only the top half of the matrix is shown and the 100% identity along the diagonal has been removed; outlined stars are used to note junction point for the row and column of the virus(es) identified. The heatmap scale in (B) is the same as found in (A). Bolded names are ICTV-recognized member species and sequences used in analysis can be found in [S2 Table](#). (C and D) Density plot of read counts per nucleotide across the viral genomes of the ourmia-like viruses (C) and mitoviruses (D) identified in this study. Graphical depiction of the genome structure, ORF, and predicted RdRP motif for each genome are along the top and a heatmap of normalized read counts is on the bottom. Read counts are normalized to a 0–1 scale, where the maximum read count for a given genome equals 1. Total reads mapped to the virus genome and average read coverage are noted. Viruses identified in this study are noted with a red star. RdRP: RNA-dependent RNA Polymerase.

(TIF)

S9 Fig. Analyses of identified viral genomes belonging to the family *Hypoviridae*. (A and C) Density plot of read counts per nucleotide across the viral genomes that group with betahypoviruses (A) and alphahypoviruses (C). Graphical depiction of the genome structure, predicted ORFs, and encoded motifs are along the top and a heatmap of normalized read counts is on the bottom. Read counts are normalized to a 0–1 scale, where the maximum read count

for a given genome equals 1. Total reads mapped to the virus genome and average read coverage are noted. (B) Percent identity matrix, generated by Clustal-Omega 2.1, of the RdRP-motif containing protein coding sequence from the newly identified viruses and related hypoviruses. For sake of clarity, the 100% identity along the diagonal has been removed and outlined stars are used to note the junction point for the row and column of the viruses identified. Where sequences have similarity $\geq 60\%$ is indicated. Bolded names are ICTV-recognized member species while italicized names are related, but unclassified viruses. Full names and accession numbers for sequences used are in [S2 Table](#). (D) Alignment of 5'UTR sequences from alphahypoviruses; conserved nucleotides are highlighted in blue and residues not included in the alignment are noted in parentheses. Viruses identified in this study are noted with a red star. Pap: papain-like cysteine protease domain homologs; RdRP: RNA-dependent RNA Polymerase; Hel: Helicase; GT: UDP-glucose/sterol glucosyltransferase; Pep: peptidase; DUF: Domain of Unknown Function.

(TIF)

S10 Fig. Analyses of viral genomes belonging with known fusariviruses. (A) Phylogram of fusariviruses based on the ORF1 protein coding sequence which contains the RdRP motif. Full names and accession numbers for protein sequences used are in [S2 Table](#), including the outgroup *Potato Virus Y* (PVY). Scale bar represent 1.0 amino acid substitutions per site and numbers at the nodes indicate bootstrap support over 70% (1000 replicates). (B) Percent identity matrix generated from Clustal 2.1 using the RdRP motif sequences from fusariviruses. For sake of clarity, the 100% identity along the diagonal has been removed and open stars note the junction point between the row and column of the viruses identified in this study. (C) Density plot of read counts per nucleotide across the viral genomes identified in this study. Graphical depiction of the genome structure, predicted ORFs, and encoded motifs are on top and a heatmap of normalized read counts is on the bottom. Read counts are normalized to a 0–1 scale, where the maximum read count for a given genome equals 1. Total reads mapped to the virus genome and average read coverage are noted. RdRP: RNA-dependent RNA Polymerase; Hel: Helicase domain; SMC: SMC N terminal domain; Spc7: Spc7 kinetochore domain.

(TIF)

S11 Fig. Sequence-based analyses of *Morchella importuna* RNA virus 1 (MiRV1). (A) Phylogram of the RdRP-motif for MiRV1, related mycoviruses, and Alpha-like viruses. Full names and accession numbers for protein sequences used are in [S2 Table](#), including the outgroup *Tobacco Mosaic Virus* (TMV). Scale bar represent 1.0 amino acid substitutions per site and numbers at the nodes indicate bootstrap support over 70% (1000 replicates). (B) Alignment of the RdRP motif from MiRV1 and related sequences. Conserved motifs within (+)ssRNA RdRP sequences are noted (I to VIII) and amino acid residues found only in the five related mycoviral sequences are highlighted in blue. Numbers in brackets are amino acids not shown. (C) Density plot of read counts per nucleotide across the MiRV1 genome. Graphical depiction of the genome structure, predicted ORF, and encoded motifs are along the top, and a heatmap of normalized read counts is on the bottom. Read counts are normalized to a 0–1 scale, where the maximum read count equals 1. Total reads mapped to the virus genome and the average read coverage are noted. (D) Percent identity matrix generated from Clustal 2.1 using the RdRP motif sequences from MiRV1, mycoviral relatives and select Alpha-like viruses. For sake of clarity, the 100% identity along the diagonal has been removed and only identities $\geq 25\%$ are indicated. Red star notes MiRV1, the virus identified in this study. RdRP: RNA-dependent RNA Polymerase; Vm: Viral methyltransferase; Hel: Helicase.

(TIF)

S1 Table. Complete list of BioProjects and SRA IDs analyzed in this study.
(XLSX)

S2 Table. Name and accession IDs of the mycoviruses used for various analyses of the viruses identified in this study, including phylogenetic trees and percent identity matrices.
Last column notes the family or genus of the mycovirus.
(XLSX)

Acknowledgments

We are deeply grateful to researchers around the world that have shared their transcriptomic data, both before and after publication. We particularly wish to acknowledge the sequencing data produced by various groups at the US Department of Energy Joint Genome Institute (<http://www.jgi.doe.gov>). We also thank the Data Science staff at the Donald Danforth Plant Science Center for excellent computational support.

Author Contributions

Conceptualization: Kerrigan B. Gilbert, James C. Carrington.

Data curation: Kerrigan B. Gilbert, Emily E. Holcomb.

Formal analysis: Kerrigan B. Gilbert, Emily E. Holcomb, Robyn L. Allscheid.

Funding acquisition: James C. Carrington.

Investigation: Emily E. Holcomb, Robyn L. Allscheid.

Project administration: Robyn L. Allscheid.

Software: Kerrigan B. Gilbert.

Supervision: Kerrigan B. Gilbert, James C. Carrington.

Visualization: Kerrigan B. Gilbert.

Writing – original draft: Kerrigan B. Gilbert, Emily E. Holcomb, James C. Carrington.

Writing – review & editing: Kerrigan B. Gilbert, Robyn L. Allscheid, James C. Carrington.

References

1. Bekal S, Domier LL, Niblack TL, Lambert KN. Discovery and initial analysis of novel viral genomes in the soybean cyst nematode. *J Gen Virol*. 2011; 92: 1870–1879. <https://doi.org/10.1099/vir.0.030585-0> PMID: [21490246](https://pubmed.ncbi.nlm.nih.gov/21490246/)
2. Mushegian A, Shipunov A, Elena SF. Changes in the composition of the RNA virome mark evolutionary transitions in green plants. *BMC Biol*. 2016; 14: 68. <https://doi.org/10.1186/s12915-016-0288-8> PMID: [27524491](https://pubmed.ncbi.nlm.nih.gov/27524491/)
3. Debat HJ. An RNA Virome Associated to the Golden Orb-Weaver Spider *Nephila clavipes*. *Front Microbiol*. 2017; 8: 2097. <https://doi.org/10.3389/fmicb.2017.02097> PMID: [29118750](https://pubmed.ncbi.nlm.nih.gov/29118750/)
4. Nerva L, Ciuffo M, Vallino M, Margaria P, Varese GC, Gnani G, et al. Multiple approaches for the detection and characterization of viral and plasmid symbionts from a collection of marine fungi. *Virus Res*. 2016; 219: 22–38. <https://doi.org/10.1016/j.virusres.2015.10.028> PMID: [26546154](https://pubmed.ncbi.nlm.nih.gov/26546154/)
5. Marzano S-YL, Nelson BD, Ajayi-Oyetunde O, Bradley CA, Hughes TJ, Hartman GL, et al. Identification of Diverse Mycoviruses through Metatranscriptomics Characterization of the Viromes of Five Major Fungal Plant Pathogens. *J Virol*. 2016; 90: 6846–6863. <https://doi.org/10.1128/JVI.00357-16> PMID: [27194764](https://pubmed.ncbi.nlm.nih.gov/27194764/)
6. Marzano S-YL, Domier LL. Novel mycoviruses discovered from metatranscriptomics survey of soybean phyllosphere phytobiomes. *Virus Res*. 2016; 213: 332–342. <https://doi.org/10.1016/j.virusres.2015.11.002> PMID: [26547008](https://pubmed.ncbi.nlm.nih.gov/26547008/)

7. Hansen JL, Long AM, Schultz SC. Structure of the RNA-dependent RNA polymerase of poliovirus. *Structure*. 1997; 5: 1109–1122. [https://doi.org/10.1016/S0969-2126\(97\)00261-X](https://doi.org/10.1016/S0969-2126(97)00261-X) PMID: 9309225
8. Campo S, Gilbert KB, Carrington JC. Small RNA-Based Antiviral Defense in the Phytopathogenic Fungus *Colletotrichum higginsianum*. *PLoS Pathog*. 2016; 12: e1005640. <https://doi.org/10.1371/journal.ppat.1005640> PMID: 27253323
9. Valero-Jiménez CA, Faino L, Spring In't Veld D, Smit S, Zwaan BJ, van Kan JAL. Comparative genomics of *Beauveria bassiana*: uncovering signatures of virulence against mosquitoes. *BMC Genomics*. 2016; 17: 986. <https://doi.org/10.1186/s12864-016-3339-1> PMID: 27905873
10. Wang J-J, Bai W-W, Zhou W, Liu J, Chen J, Liu X-Y, et al. Transcriptomic analysis of two *Beauveria bassiana* strains grown on cuticle extracts of the silkworm uncovers their different metabolic response at early infection stage. *J Invertebr Pathol*. 2017; 145: 45–54. <https://doi.org/10.1016/j.jip.2017.03.010> PMID: 28359865
11. Mondo SJ, Dannebaum RO, Kuo RC, Louie KB, Bewick AJ, LaButti K, et al. Widespread adenine N6-methylation of active genes in fungi. *Nat Genet*. 2017; 49: 964–968. <https://doi.org/10.1038/ng.3859> PMID: 28481340
12. Nerva L, Silvestri A, Ciuffo M, Palmano S, Varese GC, Turina M. Transmission of *Penicillium aurantiogriseum* partiti-like virus 1 to a new fungal host (*Cryphonectria parasitica*) confers higher resistance to salinity and reveals adaptive genomic changes. *Environ Microbiol*. 2017; 19: 4480–4492. <https://doi.org/10.1111/1462-2920.13894> PMID: 28836717
13. Liu K, Zhang W, Lai Y, Xiang M, Wang X, Zhang X, et al. *Drechslerella stenobrocha* genome illustrates the mechanism of constricting rings and the origin of nematode predation in fungi. *BMC Genomics*. 2014; 15: 114. <https://doi.org/10.1186/1471-2164-15-114> PMID: 24507587
14. Wang S, Zhang J, Li P, Qiu D, Guo L. Transcriptome-Based Discovery of *Fusarium graminearum* Stress Responses to FgHV1 Infection. *Int J Mol Sci*. 2016; 17. <https://doi.org/10.3390/ijms17111922> PMID: 27869679
15. Vanheule A, Audenaert K, Warris S, van de Geest H, Schijlen E, Höfte M, et al. Living apart together: crosstalk between the core and supernumerary genomes in a fungal plant pathogen. *BMC Genomics*. 2016; 17: 670. <https://doi.org/10.1186/s12864-016-2941-6> PMID: 27552804
16. Yang L, Xie L, Xue B, Goodwin PH, Quan X, Zheng C, et al. Comparative transcriptome profiling of the early infection of wheat roots by *Gaeumannomyces graminis* var. *tritici*. *PLoS One*. 2015; 10: e0120691. <https://doi.org/10.1371/journal.pone.0120691> PMID: 25875107
17. Wang Y, Lim L, DiGuistini S, Robertson G, Bohlmann J, Breuil C. A specialized ABC efflux transporter GcABC-G1 confers monoterpene resistance to *Grosmannia clavigera*, a bark beetle-associated fungal pathogen of pine trees. *New Phytol*. 2013; 197: 886–898. <https://doi.org/10.1111/nph.12063> PMID: 23252416
18. Sinha S, Flibotte S, Neira M, Formby S, Plemenitaš A, Cimerman NG, et al. Insight into the Recent Genome Duplication of the Halophilic Yeast *Hortaea werneckii*: Combining an Improved Genome with Gene Expression and Chromatin Structure. *G3*. 2017; 7: 2015–2022. <https://doi.org/10.1534/g3.117.040691> PMID: 28500048
19. Luo J, Qiu H, Cai G, Wagner NE, Bhattacharya D, Zhang N. Phylogenomic analysis uncovers the evolutionary history of nutrition and infection mode in rice blast fungus and other Magnaporthales. *Sci Rep*. 2015; 5: 9448. <https://doi.org/10.1038/srep09448> PMID: 25819715
20. Lehr NA, Wang Z, Li N, Hewitt DA, López-Giráldez F, Trail F, et al. Gene expression differences among three *Neurospora* species reveal genes required for sexual reproduction in *Neurospora crassa*. *PLoS One*. 2014; 9: e110398. <https://doi.org/10.1371/journal.pone.0110398> PMID: 25329823
21. Li Y, Hsiang T, Yang R-H, Hu X-D, Wang K, Wang W-J, et al. Comparison of different sequencing and assembly strategies for a repeat-rich fungal genome, *Ophiocordyceps sinensis*. *J Microbiol Methods*. 2016; 128: 1–6. <https://doi.org/10.1016/j.mimet.2016.06.025> PMID: 27343682
22. Horn F, Linde J, Mattern DJ, Walther G, Guthke R, Brakhage AA, et al. Draft Genome Sequence of the Fungus *Penicillium brasilianum* MG11. *Genome Announc*. 2015; 3. <https://doi.org/10.1128/genomeA.00724-15> PMID: 26337871
23. Wang M, Sun X, Zhu C, Xu Q, Ruan R, Yu D, et al. Pdb1A, PdabaA and PdwtA control distinct stages of conidiogenesis in *Penicillium digitatum*. *Res Microbiol*. 2015; 166: 56–65. <https://doi.org/10.1016/j.resmic.2014.12.003> PMID: 25530311
24. Knapp DG, Németh JB, Barry K, Hainaut M, Henrissat B, Johnson J, et al. Comparative genomics provides insights into the lifestyle and reveals functional heterogeneity of dark septate endophytic fungi. *Sci Rep*. 2018; 8: 6321. <https://doi.org/10.1038/s41598-018-24686-4> PMID: 29679020
25. Hulvey J, Popko JT Jr, Sang H, Berg A, Jung G. Overexpression of ShCYP51B and ShatrD in *Sclerotinia homoeocarpa* isolates exhibiting practical field resistance to a demethylation inhibitor fungicide.

- Appl Environ Microbiol. 2012; 78: 6674–6682. <https://doi.org/10.1128/AEM.00417-12> PMID: [22798361](https://pubmed.ncbi.nlm.nih.gov/22798361/)
26. Quandt CA, Di Y, Elser J, Jaiswal P, Spatafora JW. Differential Expression of Genes Involved in Host Recognition, Attachment, and Degradation in the Mycoparasite *Tolypocladium ophioglossoides*. G3. 2016; 6: 731–741. <https://doi.org/10.1534/g3.116.027045> PMID: [26801645](https://pubmed.ncbi.nlm.nih.gov/26801645/)
 27. Bech L, Busk PK, Lange L, Others. Cell wall degrading enzymes in *Trichoderma asperellum* grown on wheat bran. Fungal Genom Biol. 2015; 4: 116. Available: <http://dx.doi.org/>
 28. Lin H, Kazlauskas RJ, Travisano M. Developmental evolution facilitates rapid adaptation. Sci Rep. 2017; 7: 15891. <https://doi.org/10.1038/s41598-017-16229-0> PMID: [29162942](https://pubmed.ncbi.nlm.nih.gov/29162942/)
 29. Steindorff AS, Ramada MHS, Coelho ASG, Miller RNG, Pappas GJ Jr, Ulhoa CJ, et al. Identification of mycoparasitism-related genes against the phytopathogen *Sclerotinia sclerotiorum* through transcriptome and expression profile analysis in *Trichoderma harzianum*. BMC Genomics. 2014; 15: 204. <https://doi.org/10.1186/1471-2164-15-204> PMID: [24635846](https://pubmed.ncbi.nlm.nih.gov/24635846/)
 30. Rudd JJ, Kanyuka K, Hassani-Pak K, Derbyshire M, Andongabo A, Devonshire J, et al. Transcriptome and metabolite profiling of the infection cycle of *Zymoseptoria tritici* on wheat reveals a biphasic interaction with plant immunity involving differential pathogen chromosomal contributions and a variation on the hemibiotrophic lifestyle definition. Plant Physiol. 2015; 167: 1158–1185. <https://doi.org/10.1104/pp.114.255927> PMID: [25596183](https://pubmed.ncbi.nlm.nih.gov/25596183/)
 31. Kellner R, Bhattacharyya A, Poppe S, Hsu TY, Brem RB, Stukenbrock EH. Expression profiling of the wheat pathogen *Zymoseptoria tritici* reveals genomic patterns of transcription and host-specific regulatory programs. Genome Biol Evol. 2014; 6: 1353–1365. <https://doi.org/10.1093/gbe/evu101> PMID: [24920004](https://pubmed.ncbi.nlm.nih.gov/24920004/)
 32. Omrane S, Sghyer H, Audéon C, Lanen C, Duplaix C, Walker A-S, et al. Fungicide efflux and the MgMFS1 transporter contribute to the multidrug resistance phenotype in *Zymoseptoria tritici* field isolates. Environ Microbiol. 2015; 17: 2805–2823. <https://doi.org/10.1111/1462-2920.12781> PMID: [25627815](https://pubmed.ncbi.nlm.nih.gov/25627815/)
 33. Zhu HJ, Chen D, Zhong J, Zhang SY, Gao BD. A novel mycovirus identified from the rice false smut fungus *Ustilaginoidea virens*. Virus Genes. 2015; 51: 159–162. <https://doi.org/10.1007/s11262-015-1212-y> PMID: [26041139](https://pubmed.ncbi.nlm.nih.gov/26041139/)
 34. Koloniuk I, Hrabáková L, Petřík K. Molecular characterization of a novel amalgavirus from the entomopathogenic fungus *Beauveria bassiana*. Arch Virol. 2015; 160: 1585–1588. <https://doi.org/10.1007/s00705-015-2416-0> PMID: [25854690](https://pubmed.ncbi.nlm.nih.gov/25854690/)
 35. Kotta-Loizou I, Sipkova J, Coutts RHA. Identification and sequence determination of a novel double-stranded RNA mycovirus from the entomopathogenic fungus *Beauveria bassiana*. Arch Virol. Springer Vienna; 2015; 160: 873–875. <https://doi.org/10.1007/s00705-014-2332-8> PMID: [25577168](https://pubmed.ncbi.nlm.nih.gov/25577168/)
 36. Lin Y, Zhang H, Zhao C, Liu S, Guo L. The complete genome sequence of a novel mycovirus from *Alternaria longipes* strain HN28. Arch Virol. 2015; 160: 577–580. <https://doi.org/10.1007/s00705-014-2218-9> PMID: [25248625](https://pubmed.ncbi.nlm.nih.gov/25248625/)
 37. Aoki N, Moriyama H, Kodama M, Arie T, Teraoka T, Fukuhara T. A novel mycovirus associated with four double-stranded RNAs affects host fungal growth in *Alternaria alternata*. Virus Res. 2009; 140: 179–187. <https://doi.org/10.1016/j.virusres.2008.12.003> PMID: [19118588](https://pubmed.ncbi.nlm.nih.gov/19118588/)
 38. Kozlakidis Z, Herrero N, Ozkan S, Kanhayuwa L, Jamal A, Bhatti MF, et al. Sequence determination of a quadripartite dsRNA virus isolated from *Aspergillus foetidus*. Arch Virol. 2013; 158: 267–272. <https://doi.org/10.1007/s00705-012-1362-3> PMID: [22760661](https://pubmed.ncbi.nlm.nih.gov/22760661/)
 39. King AMQ, Lefkowitz E, Adams MJ, Carstens EB. Virus Taxonomy: Ninth Report of the International Committee on Taxonomy of Viruses [Internet]. Elsevier; 2011. Available: <https://market.android.com/details?id=book-aFYaE9KXEXUC>
 40. Baeza M, Bravo N, Sanhueza M, Flores O, Villarreal P, Cifuentes V. Molecular characterization of totiviruses in *Xanthophyllomyces dendrorhous*. Virol J. 2012; 9: 140. <https://doi.org/10.1186/1743-422X-9-140> PMID: [22838956](https://pubmed.ncbi.nlm.nih.gov/22838956/)
 41. Shi M, Lin X-D, Tian J-H, Chen L-J, Chen X, Li C-X, et al. Redefining the invertebrate RNA virosphere. Nature. 2016; <https://doi.org/10.1038/nature20167>
 42. International Committee on Taxonomy of Viruses, King AMQ. Virus Taxonomy: Classification and Nomenclature of Viruses: Ninth Report of the International Committee on Taxonomy of Viruses [Internet]. Elsevier; 2012. Available: <https://market.android.com/details?id=book-KXRCYay3pH4C>
 43. Ghabrial SA, Nibert ML. Victorivirus, a new genus of fungal viruses in the family Totiviridae. Arch Virol. 2009; 154: 373–379. <https://doi.org/10.1007/s00705-008-0272-x> PMID: [19107323](https://pubmed.ncbi.nlm.nih.gov/19107323/)

44. Komatsu K, Katayama Y, Omatsu T, Mizutani T, Fukuhara T, Kodama M, et al. Genome sequence of a novel victorivirus identified in the phytopathogenic fungus *Alternaria arborescens*. *Arch Virol*. 2016; 161: 1701–1704. <https://doi.org/10.1007/s00705-016-2796-9> PMID: 26923927
45. Romo M, Leuchtmann A, García B, Zabalgoceazcoa I. A totivirus infecting the mutualistic fungal endophyte *Epichloë festucae*. *Virus Res*. 2007; 124: 38–43. <https://doi.org/10.1016/j.virusres.2006.09.008> PMID: 17081641
46. Huang S, Ghabrial SA. Organization and expression of the double-stranded RNA genome of *Helminthosporium victoriae* 190S virus, a totivirus infecting a plant pathogenic filamentous fungus. *Proc Natl Acad Sci U S A*. 1996; 93: 12541–12546. Available: <https://www.ncbi.nlm.nih.gov/pubmed/8901618> <https://doi.org/10.1073/pnas.93.22.12541> PMID: 8901618
47. Li H, Havens WM, Nibert ML, Ghabrial SA. RNA sequence determinants of a coupled termination-reinitiation strategy for downstream open reading frame translation in *Helminthosporium victoriae* virus 190S and other victoriviruses (Family Totiviridae). *J Virol*. 2011; 85: 7343–7352. <https://doi.org/10.1128/JVI.00364-11> PMID: 21543470
48. Preisig O, Wingfield BD, Wingfield MJ. Coinfection of a fungal pathogen by two distinct double-stranded RNA viruses. *Virology*. 1998; 252: 399–406. <https://doi.org/10.1006/viro.1998.9480> PMID: 9878619
49. Niu Y, Zhang T, Zhu Y, Yuan Y, Wang S, Liu J, et al. Isolation and characterization of a novel mycovirus from *Penicillium digitatum*. *Virology*. 2016; 494: 15–22. <https://doi.org/10.1016/j.viro.2016.04.004> PMID: 27061053
50. Vainio EJ, Chiba S, Ghabrial SA, Maiss E, Roossinck M, Sabanadzovic S, et al. ICTV Virus Taxonomy Profile: Partitiviridae. *J Gen Virol*. 2018; 99: 17–18. <https://doi.org/10.1099/jgv.0.000985> PMID: 29214972
51. Covelli L, Kozlakidis Z, Di Serio F, Citir A, Açıkgöz S, Hernández C, et al. Sequences of the smallest double-stranded RNAs associated with cherry chlorotic rusty spot and Amasya cherry diseases. *Arch Virol*. 2008; 153: 759–762. <https://doi.org/10.1007/s00705-008-0039-4> PMID: 18283410
52. Osaki H, Sasaki A, Nomiyama K, Tomioka K. Multiple virus infection in a single strain of *Fusarium poae* shown by deep sequencing. *Virus Genes*. 2016; 52: 835–847. <https://doi.org/10.1007/s11262-016-1379-x> PMID: 27550368
53. Coppel P, Papp I, Bibó M, Fekete C, Hornok L. Genetic interrelationships and genome organization of double-stranded RNA elements of *Fusarium poae*. *Virus Genes*. 1999; 18: 49–56. Available: <https://www.ncbi.nlm.nih.gov/pubmed/10334037> PMID: 10334037
54. Xavier A da S, Barros APO de, Godinho MT, Zerbini FM, Souza F de O, Bruckner FP, et al. A novel mycovirus associated to *Alternaria alternata* comprises a distinct lineage in Partitiviridae. *Virus Res*. 2017; 244: 21–26. <https://doi.org/10.1016/j.virusres.2017.10.007> PMID: 29024687
55. Wang L, Jiang J, Wang Y, Hong N, Zhang F, Xu W, et al. Hypovirulence of the phytopathogenic fungus *Botryosphaeria dothidea*: association with a coinfecting chrysovirus and a partitivirus. *J Virol*. 2014; 88: 7517–7527. <https://doi.org/10.1128/JVI.00538-14> PMID: 24760881
56. Ghabrial SA, Castón JR, Coutts RHA, Hillman BI, Jiang D, Kim D-H, et al. ICTV Virus Taxonomy Profile: Chrysoviriidae. *J Gen Virol*. 2018; 99: 19–20. <https://doi.org/10.1099/jgv.0.000994> PMID: 29251589
57. Herrero N. Identification and sequence determination of a new chrysovirus infecting the entomopathogenic fungus *Isaria javanica*. *Arch Virol*. 2017; 162: 1113–1117. <https://doi.org/10.1007/s00705-016-3194-z> PMID: 28004251
58. Jiang D, Ghabrial SA. Molecular characterization of *Penicillium chrysogenum* virus: reconsideration of the taxonomy of the genus Chrysovirus. *J Gen Virol*. 2004; 85: 2111–2121. <https://doi.org/10.1099/vir.0.79842-0> PMID: 15218197
59. Covelli L, Coutts RHA, Di Serio F, Citir A, Açıkgöz S, Hernández C, et al. Cherry chlorotic rusty spot and Amasya cherry diseases are associated with a complex pattern of mycoviral-like double-stranded RNAs. I. Characterization of a new species in the genus Chrysovirus. *J Gen Virol*. 2004; 85: 3389–3397. <https://doi.org/10.1099/vir.0.80181-0> PMID: 15483256
60. Kim J-M, Park J-A, Park S-M, Cha B-J, Yang M-S, Kim D-H. Nucleotide sequences of four segments of chrysovirus in Korean *Cryphonectria nitschkei* BS122 strain. *Virus Genes*. 2010; 41: 292–294. <https://doi.org/10.1007/s11262-010-0495-2> PMID: 20549324
61. Khalifa ME, Pearson MN. Molecular characterisation of an endornavirus infecting the phytopathogen *Sclerotinia sclerotiorum*. *Virus Res*. 2014; 189: 303–309. <https://doi.org/10.1016/j.virusres.2014.06.010> PMID: 24979045
62. Moriyama H, Horiuchi H, Koga R, Fukuhara T. Molecular characterization of two endogenous double-stranded RNAs in rice and their inheritance by interspecific hybrids. *J Biol Chem*. 1999; 274: 6882–6888. <https://doi.org/10.1074/jbc.274.11.6882> PMID: 10066741

63. Ong JWL, Li H, Sivasithamparam K, Dixon KW, Jones MGK, Wylie SJ. Novel Endornavirus-like viruses, including three with two open reading frames, challenge the membership criteria and taxonomy of the Endornaviridae. *Virology*. 2016; 499: 203–211. <https://doi.org/10.1016/j.virol.2016.08.019> PMID: [27677157](https://pubmed.ncbi.nlm.nih.gov/27677157/)
64. Das S, Falloon RE, Stewart A, Pitman AR. Molecular characterisation of an endornavirus from *Rhizoctonia solani* AG-3PT infecting potato. *Fungal Biol*. 2014; 118: 924–934. <https://doi.org/10.1016/j.funbio.2014.08.003> PMID: [25442295](https://pubmed.ncbi.nlm.nih.gov/25442295/)
65. Moriyama H, Nitta T, Fukuhara T. Double-stranded RNA in rice: a novel RNA replicon in plants. *Mol Gen Genet*. 1995; 248: 364–369. Available: <https://www.ncbi.nlm.nih.gov/pubmed/7565598> PMID: [7565598](https://pubmed.ncbi.nlm.nih.gov/7565598/)
66. Pandey B, Naidu RA, Grove GG. Next generation sequencing analysis of double-stranded RNAs from sweet cherry powdery mildew fungus *Podosphaera prunicola*. *J Plant Pathol*. 2018; <https://doi.org/10.1007/s42161-018-0092-0>
67. Adams MJ, Adkins S, Bragard C, Gilmer D, Li D, MacFarlane SA, et al. ICTV Virus Taxonomy Profile: Virgaviridae. *J Gen Virol*. 2017; 98: 1999–2000. <https://doi.org/10.1099/jgv.0.000884> PMID: [28786782](https://pubmed.ncbi.nlm.nih.gov/28786782/)
68. Mascia T, Nigro F, Abdallah A, Ferrara M, De Stradis A, Faedda R, et al. Gene silencing and gene expression in phytopathogenic fungi using a plant virus vector. *Proc Natl Acad Sci U S A*. 2014; 111: 4291–4296. <https://doi.org/10.1073/pnas.1315668111> PMID: [24594602](https://pubmed.ncbi.nlm.nih.gov/24594602/)
69. Andika IB, Wei S, Cao C, Salaipeth L, Kondo H, Sun L. Phytopathogenic fungus hosts a plant virus: A naturally occurring cross-kingdom viral infection. *Proc Natl Acad Sci U S A*. 2017; 114: 12267–12272. <https://doi.org/10.1073/pnas.1714916114> PMID: [29087346](https://pubmed.ncbi.nlm.nih.gov/29087346/)
70. Mosier AC, Miller CS, Frischkorn KR, Ohm RA, Li Z, LaButti K, et al. Fungi Contribute Critical but Spatially Varying Roles in Nitrogen and Carbon Cycling in Acid Mine Drainage. *Front Microbiol*. 2016; 7: 238. <https://doi.org/10.3389/fmicb.2016.00238> PMID: [26973616](https://pubmed.ncbi.nlm.nih.gov/26973616/)
71. Deneff VJ, Mueller RS, Banfield JF. AMD biofilms: using model communities to study microbial evolution and ecological complexity in nature. *ISME J*. 2010; 4: 599–610. <https://doi.org/10.1038/ismej.2009.158> PMID: [20164865](https://pubmed.ncbi.nlm.nih.gov/20164865/)
72. Andersson AF, Banfield JF. Virus population dynamics and acquired virus resistance in natural microbial communities. *Science*. 2008; 320: 1047–1050. <https://doi.org/10.1126/science.1157358> PMID: [18497291](https://pubmed.ncbi.nlm.nih.gov/18497291/)
73. Cañizares MC, López-Escudero FJ, Pérez-Artés E, García-Pedrajas MD. Characterization of a novel single-stranded RNA mycovirus related to invertebrate viruses from the plant pathogen *Verticillium dahliae*. *Arch Virol*. 2018; 163: 771–776. <https://doi.org/10.1007/s00705-017-3644-2> PMID: [29147792](https://pubmed.ncbi.nlm.nih.gov/29147792/)
74. Ai Y-P, Zhong J, Chen C-Y, Zhu H-J, Gao B-D. A novel single-stranded RNA virus isolated from the rice-pathogenic fungus *Magnaporthe oryzae* with similarity to members of the family Tombusviridae. *Arch Virol*. 2016; 161: 725–729. <https://doi.org/10.1007/s00705-015-2683-9> PMID: [26650038](https://pubmed.ncbi.nlm.nih.gov/26650038/)
75. Preisig O, Moleleki N, Smit WA, Wingfield BD, Wingfield MJ. A novel RNA mycovirus in a hypovirulent isolate of the plant pathogen *Diaporthe ambigua*. *J Gen Virol*. 2000; 81: 3107–3114. <https://doi.org/10.1099/0022-1317-81-12-3107> PMID: [11086142](https://pubmed.ncbi.nlm.nih.gov/11086142/)
76. Poch O, Sauvaget I, Delarue M, Tordo N. Identification of four conserved motifs among the RNA-dependent polymerase encoding elements. *EMBO J*. 1989; 8: 3867–3874. Available: <https://www.ncbi.nlm.nih.gov/pubmed/2555175> PMID: [2555175](https://pubmed.ncbi.nlm.nih.gov/2555175/)
77. Vázquez AL, Alonso JM, Parra F. Mutation analysis of the GDD sequence motif of a calicivirus RNA-dependent RNA polymerase. *J Virol*. 2000; 74: 3888–3891. Available: <https://www.ncbi.nlm.nih.gov/pubmed/10729164> <https://doi.org/10.1128/jvi.74.8.3888-3891.2000> PMID: [10729164](https://pubmed.ncbi.nlm.nih.gov/10729164/)
78. Lohmann V, Körner F, Herian U, Bartenschlager R. Biochemical properties of hepatitis C virus NS5B RNA-dependent RNA polymerase and identification of amino acid sequence motifs essential for enzymatic activity. *J Virol*. 1997; 71: 8416–8428. Available: <https://www.ncbi.nlm.nih.gov/pubmed/9343198> PMID: [9343198](https://pubmed.ncbi.nlm.nih.gov/9343198/)
79. Jablonski SA, Morrow CD. Mutation of the aspartic acid residues of the GDD sequence motif of poliovirus RNA-dependent RNA polymerase results in enzymes with altered metal ion requirements for activity. *J Virol*. 1995; 69: 1532–1539. Available: <https://www.ncbi.nlm.nih.gov/pubmed/7853486> PMID: [7853486](https://pubmed.ncbi.nlm.nih.gov/7853486/)
80. O'Reilly EK, Kao CC. Analysis of RNA-dependent RNA polymerase structure and function as guided by known polymerase structures and computer predictions of secondary structure. *Virology*. 1998; 252: 287–303. <https://doi.org/10.1006/viro.1998.9463> PMID: [9878607](https://pubmed.ncbi.nlm.nih.gov/9878607/)
81. Moleleki N, van Heerden SW, Wingfield MJ, Wingfield BD, Preisig O. Transfection of *Diaporthe perijuncta* with *Diaporthe* RNA virus. *Appl Environ Microbiol*. 2003; 69: 3952–3956. Available: <https://doi.org/10.1128/AEM.69.11.3952-3956.2003> PMID: [12841111](https://pubmed.ncbi.nlm.nih.gov/12841111/)

- www.ncbi.nlm.nih.gov/pubmed/12839766 <https://doi.org/10.1128/AEM.69.7.3952-3956.2003> PMID: 12839766
82. Hisano S, Zhang R, Faruk MI, Kondo H, Suzuki N. A neo-virus lifestyle exhibited by a (+)ssRNA virus hosted in an unrelated dsRNA virus: Taxonomic and evolutionary considerations. *Virus Res.* 2018; 244: 75–83. <https://doi.org/10.1016/j.virusres.2017.11.006> PMID: 29122644
 83. Kozlakidis Z, Herrero N, Ozkan S, Bhatti MF, Coutts RHA. A novel dsRNA element isolated from the *Aspergillus foetidus* mycovirus complex. *Arch Virol.* 2013; 158: 2625–2628. <https://doi.org/10.1007/s00705-013-1779-3> PMID: 23827976
 84. Zhang R, Hisano S, Tani A, Kondo H, Kanematsu S, Suzuki N. A capsidless ssRNA virus hosted by an unrelated dsRNA virus. *Nat Microbiol.* 2016; 1: 15001. <https://doi.org/10.1038/nmicrobiol.2015.1> PMID: 27571749
 85. Koonin EV, Wolf YI, Nagasaki K, Dolja VV. The Big Bang of picorna-like virus evolution antedates the radiation of eukaryotic supergroups. *Nat Rev Microbiol.* 2008; 6: 925–939. <https://doi.org/10.1038/nrmicro2030> PMID: 18997823
 86. Kotta-Loizou I, Coutts RHA. Mycoviruses in *Aspergilli*: A Comprehensive Review. *Front Microbiol.* 2017; 8: 1699. <https://doi.org/10.3389/fmicb.2017.01699> PMID: 28932216
 87. Hillman BI, Cai G. The family *narnaviridae*: simplest of RNA viruses. *Adv Virus Res.* 2013; 86: 149–176. <https://doi.org/10.1016/B978-0-12-394315-6.00006-4> PMID: 23498906
 88. Osawa S, Jukes TH, Watanabe K, Muto A. Recent evidence for evolution of the genetic code. *Microbiol Rev.* 1992; 56: 229–264. Available: <https://www.ncbi.nlm.nih.gov/pubmed/1579111> PMID: 1579111
 89. Ashwin NMR, Barnabas L, Ramesh Sundar A, Malathi P, Viswanathan R, Masi A, et al. Comparative secretome analysis of *Colletotrichum falcatum* identifies a cerato-platanin protein (EPL1) as a potential pathogen-associated molecular pattern (PAMP) inducing systemic resistance in sugarcane. *J Proteomics.* 2017; 169: 2–20. <https://doi.org/10.1016/j.jprot.2017.05.020> PMID: 28546091
 90. Turina M, Hillman BI, Izadpanah K, Rastgou M, Rosa C, Ictv Report Consortium. ICTV Virus Taxonomy Profile: Ourmiavirus. *J Gen Virol.* 2017; 98: 129–130. <https://doi.org/10.1099/jgv.0.000725> PMID: 28284246
 91. Hrabáková L, Koloniuk I, Petrzik K. *Phomopsis longicolla* RNA virus 1 - Novel virus at the edge of myco- and plant viruses. *Virology.* 2017; 506: 14–18. <https://doi.org/10.1016/j.virol.2017.03.003> PMID: 28288321
 92. Donaire L, Rozas J, Ayllón MA. Molecular characterization of *Botrytis ourmia*-like virus, a mycovirus close to the plant pathogenic genus *Ourmiavirus*. *Virology.* 2016; 489: 158–164. <https://doi.org/10.1016/j.virol.2015.11.027> PMID: 26765967
 93. Suzuki N, Ghabrial SA, Kim K-H, Pearson M, Marzano S-YL, Yaegashi H, et al. ICTV Virus Taxonomy Profile: Hypoviridae. *J Gen Virol.* 2018; 99: 615–616. <https://doi.org/10.1099/jgv.0.001055> PMID: 29589826
 94. Fahima T, Kazmierczak P, Hansen DR, Pfeiffer P, Van Alfen NK. Membrane-associated replication of an unencapsidated double-strand RNA of the fungus, *Cryphonectria parasitica*. *Virology.* 1993; 195: 81–89. <https://doi.org/10.1006/viro.1993.1348> PMID: 8317109
 95. Jacob-Wilk D, Turina M, Van Alfen NK. Mycovirus *Cryphonectria hypovirus 1* elements cofractionate with trans-Golgi network membranes of the fungal host *Cryphonectria parasitica*. *J Virol.* 2006; 80: 6588–6596. <https://doi.org/10.1128/JVI.02519-05> PMID: 16775345
 96. Li P, Zhang H, Chen X, Qiu D, Guo L. Molecular characterization of a novel hypovirus from the plant pathogenic fungus *Fusarium graminearum*. *Virology.* 2015; 481: 151–160. <https://doi.org/10.1016/j.virol.2015.02.047> PMID: 25781585
 97. Yaegashi H, Kanematsu S, Ito T. Molecular characterization of a new hypovirus infecting a phytopathogenic fungus, *Valsa ceratosperma*. *Virus Res.* 2012; 165: 143–150. <https://doi.org/10.1016/j.virusres.2012.02.008> PMID: 22366520
 98. Shapira R, Choi GH, Nuss DL. Virus-like genetic organization and expression strategy for a double-stranded RNA genetic element associated with biological control of chestnut blight. *EMBO J.* 1991; 10: 731–739. Available: <https://www.ncbi.nlm.nih.gov/pubmed/2009854> PMID: 2009854
 99. Hillman BI, Halpern BT, Brown MP. A viral dsRNA element of the chestnut blight fungus with a distinct genetic organization. *Virology.* 1994; 201: 241–250. <https://doi.org/10.1006/viro.1994.1289> PMID: 8184535
 100. Arjona-Lopez JM, Telengech P, Jamal A, Hisano S, Kondo H, Yelin MD, et al. Novel, diverse RNA viruses from Mediterranean isolates of the phytopathogenic fungus, *Rosellinia necatrix*: insights into evolutionary biology of fungal viruses. *Environ Microbiol.* 2018; 20: 1464–1483. <https://doi.org/10.1111/1462-2920.14065> PMID: 29411500

101. Li P, Chen X, He H, Qiu D, Guo L. Complete Genome Sequence of a Novel Hypovirus from the Phytopathogenic Fungus *Fusarium langsethiae*. *Genome Announc.* 2017; 5. <https://doi.org/10.1128/genomeA.01722-16> PMID: 28254984
102. Zhang R, Liu S, Chiba S, Kondo H, Kanematsu S, Suzuki N. A novel single-stranded RNA virus isolated from a phytopathogenic filamentous fungus, *Rosellinia necatrix*, with similarity to hypo-like viruses. *Front Microbiol.* 2014; 5: 360. <https://doi.org/10.3389/fmicb.2014.00360> PMID: 25101066
103. Hrabáková L, Grum-Grzhimaylo AA, Koloniuk I, Debets AJM, Sarkisova T, Petrzik K. The alkalophilic fungus *Sodiomyces alkalinus* hosts beta- and gammapartitiviruses together with a new fusarivirus. *PLoS One.* 2017; 12: e0187799. <https://doi.org/10.1371/journal.pone.0187799> PMID: 29186149
104. Liu R, Cheng J, Fu Y, Jiang D, Xie J. Molecular Characterization of a Novel Positive-Sense, Single-Stranded RNA Mycovirus Infecting the Plant Pathogenic Fungus *Sclerotinia sclerotiorum*. *Viruses.* 2015; 7: 2470–2484. <https://doi.org/10.3390/v7052470> PMID: 26008696
105. Liu H, Fu Y, Jiang D, Li G, Xie J, Peng Y, et al. A novel mycovirus that is related to the human pathogen hepatitis E virus and rubi-like viruses. *J Virol.* 2009; 83: 1981–1991. <https://doi.org/10.1128/JVI.01897-08> PMID: 19073734
106. Bartholomäus A, Wibberg D, Winkler A, Pühler A, Schlüter A, Varrelmann M. Deep Sequencing Analysis Reveals the Mycoviral Diversity of the Virome of an Avirulent Isolate of *Rhizoctonia solani* AG-2-2 IV. *PLoS One.* 2016; 11: e0165965. <https://doi.org/10.1371/journal.pone.0165965> PMID: 27814394
107. Koonin EV, Dolja VV, Krupovic M. Origins and evolution of viruses of eukaryotes: The ultimate modularity. *Virology.* 2015; 479–480: 2–25. <https://doi.org/10.1016/j.virol.2015.02.039> PMID: 25771806
108. Hacquard S, Kracher B, Hiruma K, Münch PC, Garrido-Oter R, Thon MR, et al. Survival trade-offs in plant roots during colonization by closely related beneficial and pathogenic fungi. *Nat Commun.* 2016; 7: 11362. <https://doi.org/10.1038/ncomms11362> PMID: 27150427
109. Grandaubert J, Bhattacharyya A, Stukenbrock EH. RNA-seq-Based Gene Annotation and Comparative Genomics of Four Fungal Grass Pathogens in the Genus *Zymoseptoria* Identify Novel Orphan Genes and Species-Specific Invasions of Transposable Elements. *G3.* 2015; 5: 1323–1333. <https://doi.org/10.1534/g3.115.017731> PMID: 25917918
110. Rajakaruna P, Khandekar S, Meulia T, Leisner SM. Identification and Host Relations of Turnip ringspot virus, A Novel Comovirus from Ohio. *Plant Dis. Scientific Societies;* 2007; 91: 1212–1220. <https://doi.org/10.1094/PDIS-91-10-1212> PMID: 30780511
111. Roossinck MJ. The good viruses: viral mutualistic symbioses. *Nat Rev Microbiol.* 2011; 9: 99–108. <https://doi.org/10.1038/nrmicro2491> PMID: 21200397
112. Schmitt MJ, Breinig F. The viral killer system in yeast: from molecular biology to application. *FEMS Microbiol Rev.* 2002; 26: 257–276. Available: <https://www.ncbi.nlm.nih.gov/pubmed/12165427> <https://doi.org/10.1111/j.1574-6976.2002.tb00614.x> PMID: 12165427
113. Márquez LM, Redman RS, Rodriguez RJ, Roossinck MJ. A virus in a fungus in a plant: three-way symbiosis required for thermal tolerance. *Science.* 2007; 315: 513–515. <https://doi.org/10.1126/science.1136237> PMID: 17255511
114. Niu Y, Yuan Y, Mao J, Yang Z, Cao Q, Zhang T, et al. Characterization of two novel mycoviruses from *Penicillium digitatum* and the related fungicide resistance analysis. *Sci Rep.* 2018; 8: 5513. <https://doi.org/10.1038/s41598-018-23807-3> PMID: 29615698
115. Chiba S, Suzuki N. Highly activated RNA silencing via strong induction of dicer by one virus can interfere with the replication of an unrelated virus. *Proc Natl Acad Sci U S A.* 2015; 112: E4911–8. <https://doi.org/10.1073/pnas.1509151112> PMID: 26283371
116. Segers GC, Zhang X, Deng F, Sun Q, Nuss DL. Evidence that RNA silencing functions as an antiviral defense mechanism in fungi. *Proc Natl Acad Sci U S A.* 2007; 104: 12902–12906. <https://doi.org/10.1073/pnas.0702500104> PMID: 17646660
117. Mochama P, Jadhav P, Neupane A, Lee Marzano S-Y. Mycoviruses as Triggers and Targets of RNA Silencing in White Mold Fungus *Sclerotinia sclerotiorum*. *Viruses.* 2018; 10. <https://doi.org/10.3390/v10040214> PMID: 29690568
118. Kirk PM, Cannon PF, David JC, Stalpers JA, editor. *Ainsworth and Bisby's Dictionary of Fungi.* CABI Bioscience; 2001.
119. Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. *Nat Methods.* 2012; 9: 357–359. <https://doi.org/10.1038/nmeth.1923> PMID: 22388286
120. Haas BJ, Papanicolaou A, Yassour M, Grabherr M, Blood PD, Bowden J, et al. De novo transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. *Nat Protoc.* 2013; 8: 1494–1512. <https://doi.org/10.1038/nprot.2013.084> PMID: 23845962
121. TransDecoder WebCitation [Internet]. [cited 16 Aug 2017]. Available: transdecoder.github.io
122. HMMER [Internet]. [cited 14 Jun 2018]. Available: <http://hmmer.org>

123. Martin M. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet journal*. 2011; 17: 10–12. <https://doi.org/10.14806/ej.17.1.200>
124. Home - ORFfinder - NCBI [Internet]. [cited 20 Dec 2017]. Available: <https://www.ncbi.nlm.nih.gov/orffinder/>
125. SankeyMATIC (BETA): A Sankey diagram builder for everyone [Internet]. [cited 14 Jun 2018]. Available: <http://sankeymatic.com/>
126. Sperschneider J, Datta A. DotKnot: pseudoknot prediction using the probability dot plot under a refined energy model. *Nucleic Acids Res*. 2010; 38: e103. <https://doi.org/10.1093/nar/gkq021> PMID: [20123730](https://pubmed.ncbi.nlm.nih.gov/20123730/)
127. Sperschneider J, Datta A, Wise MJ. Heuristic RNA pseudoknot prediction including intramolecular kissing hairpins. *RNA*. 2011; 17: 27–38. <https://doi.org/10.1261/rna.2394511> PMID: [21098139](https://pubmed.ncbi.nlm.nih.gov/21098139/)
128. Byun Y, Han K. PseudoViewer3: generating planar drawings of large-scale RNA structures with pseudoknots. *Bioinformatics*. 2009; 25: 1435–1437. <https://doi.org/10.1093/bioinformatics/btp252> PMID: [19369500](https://pubmed.ncbi.nlm.nih.gov/19369500/)
129. Katoh K, Standley DM. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol Biol Evol*. 2013; 30: 772–780. <https://doi.org/10.1093/molbev/mst010> PMID: [23329690](https://pubmed.ncbi.nlm.nih.gov/23329690/)
130. Capella-Gutiérrez S, Silla-Martínez JM, Gabaldón T. trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics*. 2009; 25: 1972–1973. <https://doi.org/10.1093/bioinformatics/btp348> PMID: [19505945](https://pubmed.ncbi.nlm.nih.gov/19505945/)
131. Stamatakis A. RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. *Bioinformatics*. 2006; 22: 2688–2690. <https://doi.org/10.1093/bioinformatics/btl446> PMID: [16928733](https://pubmed.ncbi.nlm.nih.gov/16928733/)
132. Huson DH, Scornavacca C. Dendroscope 3: an interactive tool for rooted phylogenetic trees and networks. *Syst Biol*. 2012; 61: 1061–1067. <https://doi.org/10.1093/sysbio/sys062> PMID: [22780991](https://pubmed.ncbi.nlm.nih.gov/22780991/)
133. Li W, Cowley A, Uludag M, Gur T, McWilliam H, Squizzato S, et al. The EMBL-EBI bioinformatics web and programmatic tools framework. *Nucleic Acids Res*. 2015; 43: W580–4. <https://doi.org/10.1093/nar/gkv279> PMID: [25845596](https://pubmed.ncbi.nlm.nih.gov/25845596/)