



## OPEN ACCESS

## EDITED BY

Richard John Philip Brown,  
Paul Ehrlich Institute, Germany

## REVIEWED BY

Daniel Todt,  
Ruhr University Bochum, Germany  
Ingrida Olendraite,  
University of Cambridge, United Kingdom

## \*CORRESPONDENCE

Chris Lauber  
✉ [chris.lauber@twincore.de](mailto:chris.lauber@twincore.de)

RECEIVED 13 January 2023

ACCEPTED 24 April 2023

PUBLISHED 12 May 2023

## CITATION

Chong LC and Lauber C (2023) Viroid-like RNA-dependent RNA polymerase-encoding ambiviruses are abundant in complex fungi. *Front. Microbiol.* 14:1144003. doi: 10.3389/fmicb.2023.1144003

## COPYRIGHT

© 2023 Chong and Lauber. This is an open-access article distributed under the terms of the [Creative Commons Attribution License \(CC BY\)](https://creativecommons.org/licenses/by/4.0/). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

# Viroid-like RNA-dependent RNA polymerase-encoding ambiviruses are abundant in complex fungi

Li Chuin Chong and Chris Lauber\*

Institute for Experimental Virology, TWINCORE Centre for Experimental and Clinical Infection Research, a Joint Venture Between the Hannover Medical School (MHH) and the Helmholtz Centre for Infection Research (HZI), Hannover, Germany

Ambiviruses are hybrid infectious elements encoding the hallmark gene of RNA viruses, the RNA-dependent RNA polymerase, and self-cleaving RNA ribozymes found in many viroids. Ambiviruses are thought to be pathogens of fungi, although the majority of reported genomes have been identified in metatranscriptomes. Here, we present a comprehensive screen for ambiviruses in more than 46,500 fungal transcriptomes from the Sequence Read Archive (SRA). Our data-driven virus discovery approach identified more than 2,500 ambiviral sequences across the kingdom *Fungi* with a striking expansion in members of the phylum *Basidiomycota* representing the most complex fungal organisms. Our study unveils a large diversity of unknown ambiviruses with as little as 27% protein sequence identity to known members and sheds new light on the evolution of this distinct class of infectious agents with RNA genomes. No evidence for the presence of ambiviruses in human microbiomes was obtained from a comprehensive screen of respective metatranscriptomes available in the SRA.

## KEYWORDS

virus discovery, ambiviruses, human metatranscriptome, computational virology, viroid-like elements, fungal pathogen

## Introduction

Infectious genetic elements with RNA genomes (IGERs) encompass viroids and viroid-like RNAs as well as certain viruses. These agents have been demonstrated to cause a multitude of economically and medically important diseases despite major differences in genome size, genetic complexity, and life cycle between the various IGER classes and subclasses. The vast diversity of known and potentially unknown IGERs makes them ideal systems to study micro- and macro-evolutionary processes, the emergence of hybrid elements with features from difference IGER classes, and the origin(s) of life.

Viroids are the simplest form among known infectious pathogens, consisting of a single-stranded, covalently closed circular RNA genome of a few hundred nucleotides in length (Diener, 2001; Daròs et al., 2006). Although the known viroids do not encode proteins, they interact with the host via RNA structures to hijack the host transcription machinery for replication of their small RNA genomes. The members of different viroid families adopt distinct RNA structures, including branched as well as rod-shaped conformations (Giguère et al., 2014). Additional viroid-like IGERs include retroviroids, which integrate into the host genome with the help of a pararetrovirus (Daròs and Flores, 1995), circular satellite RNAs of plants, which require a helper virus for their replication and transmission (Bruening et al., 1991; Rao and Kalantidis, 2015), and animal-infecting ribozyviruses, including the important human pathogen

hepatitis delta virus (HDV). Many viroids and viroid-like IGERs utilize IGER-encoded ribozymes, such as hammerhead ribozyme (HHR) or hairpin ribozyme (HPR), for cleavage of their multimeric replication products (Kos et al., 1986; Branch et al., 1988; Modahl et al., 2000; Ferre-D'Amare and Scott, 2010; Sureau and Negro, 2016; de la Peña et al., 2020; Wang, 2021). Known members of these classes have RNA genomes well below 2,000 nt.

Recently described ambiviruses employ considerably larger genomes in the range between 4,000 and 5,000 nt (Sutela et al., 2020; Forgia et al., 2021). Their circular RNA genomes exhibit unique features that make them hybrids between RNA viruses and viroids: they have two open reading frames (ORFs), one of which encodes an RNA-dependent RNA polymerase (RdRp) related to the RdRps of RNA viruses, while the function of the product encoded by the second ORF remains unknown. In addition, ambiviruses code for two HHR, HPR, or other types of ribozymes that are located close to each ORF's C-terminal part in the non-protein-coding region of the genome. The two ORF-ribozyme pairs are encoded on opposite genome polarities. Ambiviruses are thought to infect fungi, although the large majority of ambivirus genomes have been discovered from metatranscriptomes (Sutela et al., 2020; Forgia et al., 2021, 2022). Due to the latter, a comprehensive and detailed ambivirus host distribution within and potentially also outside the kingdom *Fungi* is lacking. This paucity includes a description of the presence or absence of ambiviruses in the human mycobiome formed by fungal components of the microbiome which interact with both the bacterial microbiome and host immunity, and can influence pathophysiological processes in humans (Nguyen et al., 2015; Soret et al., 2020; Pérez, 2021; Zhang et al., 2022).

Here, we report the results of a screen for ambivirus genomes in almost 60,000 transcriptome projects of fungi and human microbiomes representing the full diversity available by the time of writing in the Sequence Read Archive (SRA). We discovered more than 2,500 viral sequences from 345 distinct ambiviruses and demonstrate an expansion of ambiviruses in the most complex fungal organisms from the phylum *Basidiomycota*. In general, our study offers new insights into the diversity and evolution of this distinct class of infectious agents.

## Results and discussion

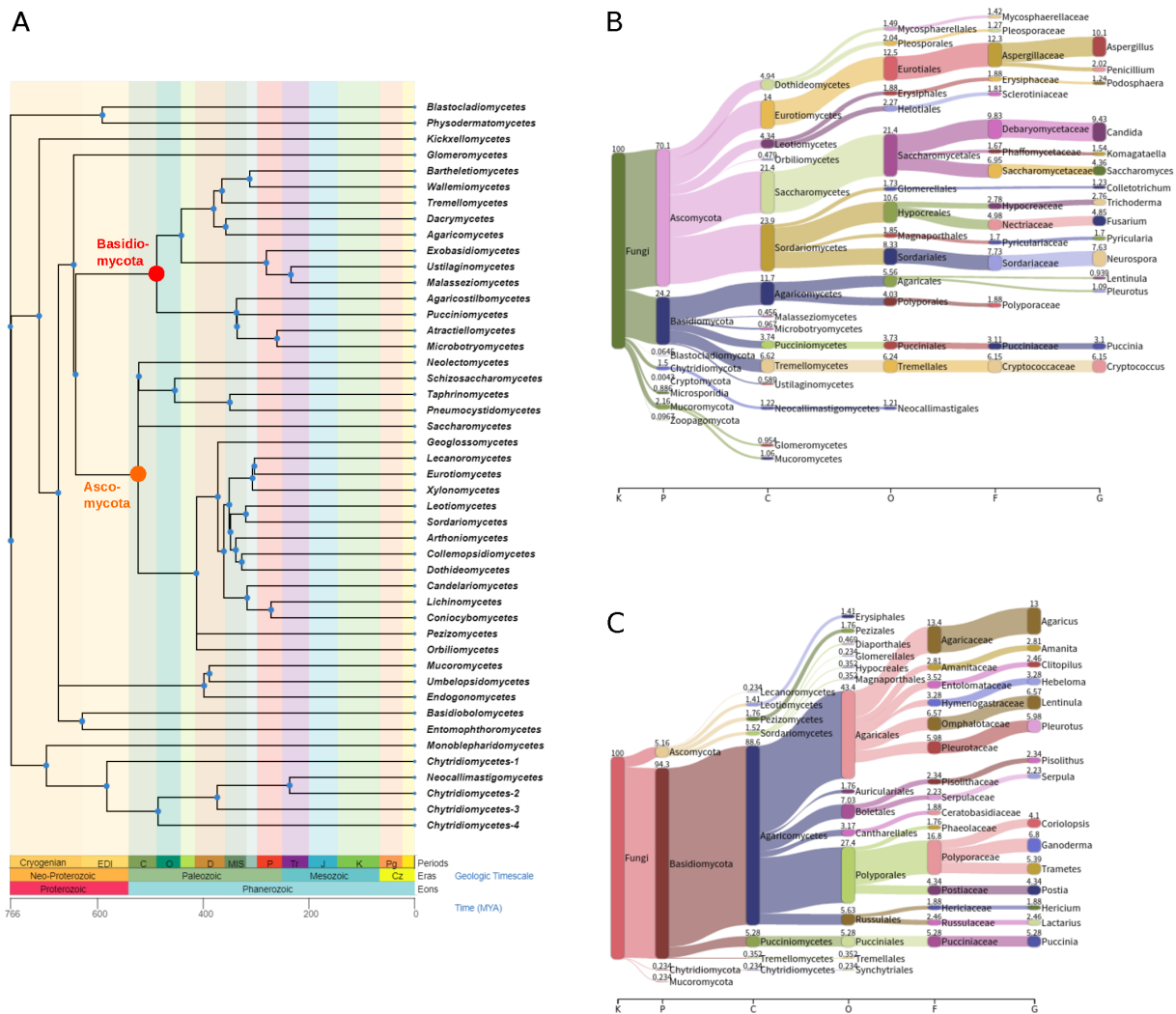
We have applied a data-driven virus discovery (DDVD) approach (Lauber and Seitz, 2022) to screen a comprehensive set of 46,519 transcriptomes from the SRA representing the global diversity of fungi (Figure 1A) for the presence of ambivirus sequences. Our screening involved a sensitive sequence homology search in raw sequencing read data using a profile Hidden Markov Model (pHMM) of the ambivirus RdRp, which obtained hits against this ambivirus-specific pHMM in 853 SRA data sets. The subsequent seed-based genome assembly specifically targeted the sequences identified in the first stage and produced 2,588 contigs with significant sequence similarity to the ambiviral RdRp region including the well-conserved motifs A, B, and C (Gorbalenya et al., 2002; Bruenn, 2003). Removal of sequence redundancy by clustering the contigs at 90% nucleotide sequence identity and RdRp fragments shorter than 500 nt resulted in 345 unique ambivirus sequences of which 81 were full-length circular RNA genomes while the other assemblies represented incomplete genomes (Figure 2A; Supplementary Table S1). They have been

retrieved from only 181 BioSamples in total (Supplementary Table S2), demonstrating the concurrent infection of individual fungi by several viruses. Although the viral reads typically constituted a minor fraction of the total amount of reads in a sequencing experiment (0.03% on average), the read depth per viral genome position was moderate to very high (273.7 on average; Figure 2A). The 345 discovered ambiviruses showed protein sequence identities to ambiviruses described in Forgia et al. (2022) and reference databases of 47.7% on average (range of 26.8–100%; Figure 2A), indicating that the majority of the ambiviruses discovered in this study are novel and that previous searches based on metatranscriptome analyses only revealed a fraction of the natural ambivirus diversity.

A strength of the SRA-based virus discovery approach is the availability of often detailed metadata, including host taxonomic information, for many of the underlying sequencing projects. We utilized this information (Supplementary Table S2) and mapped the fungal taxonomy of each sequencing project to the ambivirus sequences discovered from this project. Strikingly, although more than 70% of the analyzed SRA experiments studied fungi were from the phylum *Ascomycota* (Figure 1B), only very few ambiviruses (5.2%) were found in these fungi (Figure 1C). In sharp contrast, the large majority (94.3%) of the discovered ambiviruses were from the phylum *Basidiomycota*, which constituted only 24.2% of the analyzed SRA data sets. The difference between *Basidiomycota* and *Ascomycota* (Fisher's exact test,  $p=0$ ) as well as *Mucoromycota* ( $p=4.1e-28$ ) and *Chytridiomycota* ( $p=3.6e-19$ ) was statistically highly significant, while no significant differences were observed between the other pairs of phyla (Supplementary Table S3).

*Ascomycota* species, including the model organism *Saccharomyces cerevisiae*, commonly (but not exclusively) reproduce asexually and are characterized by internal spore production in a sac-like structure called the ascus. Members from the *Basidiomycota* form spores externally by specialized cells called basidia, and sexual reproduction is considered to be more common among *Basidiomycota* species. It is tempting to speculate that the mode of reproduction may play a role in the spread of ambiviruses, a hypothesis that warrants further investigation, for instance via comparative infection experiments. Another factor of susceptibility to ambivirus infection might be linked to the higher complexity, in terms of cell cycle and multicellularity, of *Basidiomycota* species compared to those of other orders (Naranjo-Ortiz and Gabaldón, 2019). *Ascomycota* and *Basidiomycota* form two sister clades in the fungal tree of life (together building the most species-rich fungal subkingdom *Dikarya*) and constitute two relatively young lineages compared to the other fungal orders (Naranjo-Ortiz and Gabaldón, 2019), indicating that the observed expansion of ambiviruses in *Basidiomycota* was established after the split of the two orders.

The novel ambiviruses with full-length or near full-length genome sequences showed the expected genomic organization involving two open reading frames (ORFs) encoded in opposite reading directions (sense and antisense; Figures 2B,C). A self-cleaving hammerhead or hairpin ribozyme was found to be encoded near the C-terminus of each of the two ORFs (Figures 2B,C). We identified structural RNA motifs in 178 of the 345 ambivirus genomic sequences. When considering the top two hits per contig, the most frequent RNA structural motif was Hammerhead\_3 ribozyme ( $n=81$ , Rfam accession: RF00008), followed by Hairpin ribozyme ( $n=77$ , RF00173) and Hairpin-meta1 ribozyme ( $n=76$ , RF04190). Similar to other viroid-like elements, such as HDV, many of the ambivirus genomes are predicted to adopt a rod-shaped



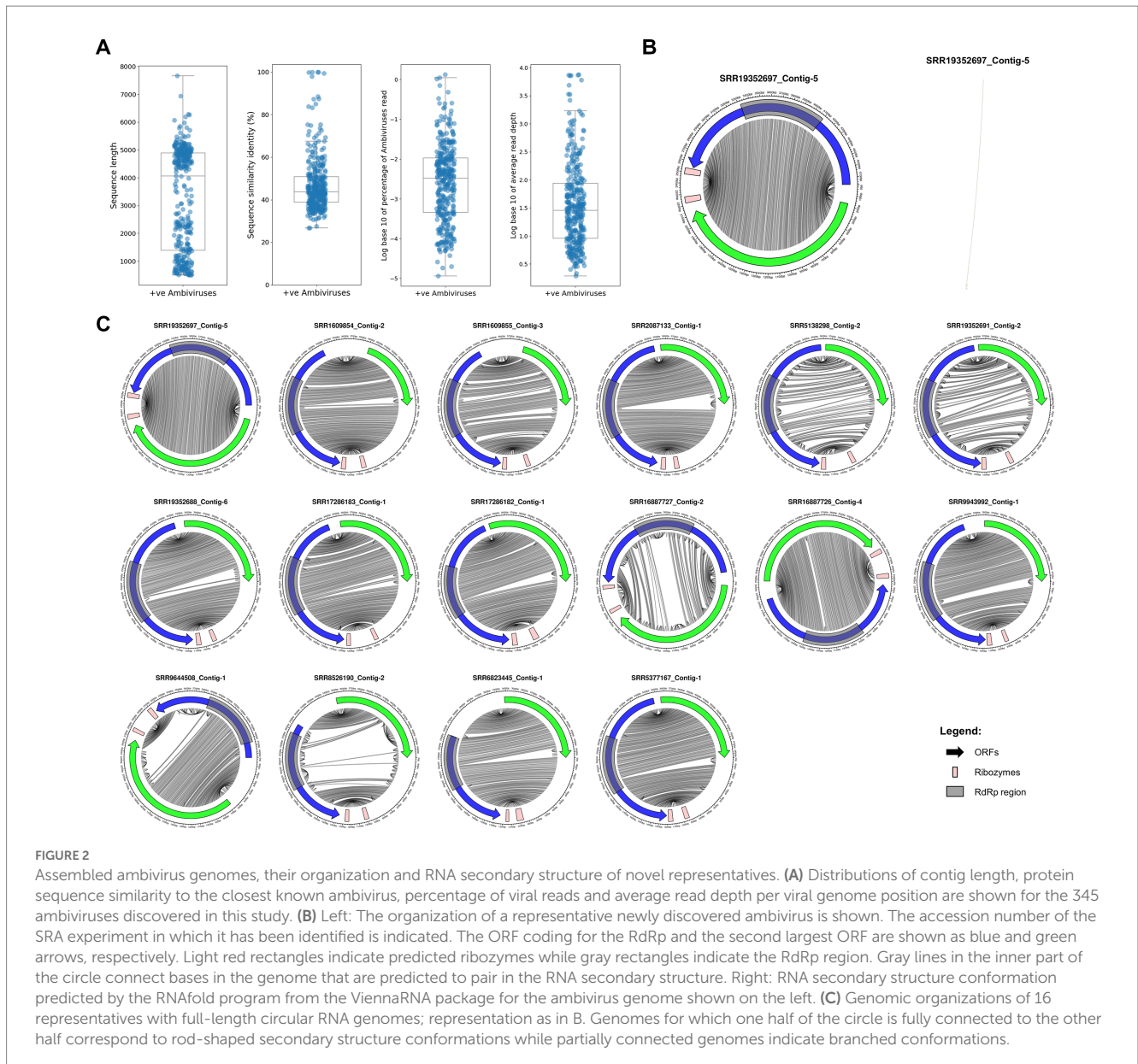
**FIGURE 1** Phylogeny of fungi and taxonomic classification of fungal transcriptome experiments from the SRA. (A) The tree shows phylogenetic relationships of fungal classes and was obtained from Timetree 5 (<http://timetree.org/>). Branch lengths are in millions of years ago (MYA). The roots of the clades for Ascomycota and Basidiomycota are highlighted in orange and red, respectively. Sankey diagrams are shown for all 46,519 fungal SRA data sets that have been analyzed (B) and for those 853 of them in which ambivirus sequences have been discovered (C). Numbers above colored bars show the percentage of SRA data sets at each taxonomic rank. Note the difference between relative frequencies of the phyla Ascomycota and Basidiomycota between the two diagrams. Taxonomic ranks shown are kingdom (K), phylum (P), class (C), order (O), family (F), and genus (G).

RNA secondary structure conformation (Figures 2B,C), while others show a branched conformation (Figure 2C).

We used RdRp protein sequences of previously described and newly discovered ambiviruses to reconstruct an ambivirus phylogeny (Figure 3). Viral groups of relatively low diversity were associated with specific fungal orders while the viral relationships predicted frequent cross-species transmissions at the macroevolutionary scale, as indicated by ambiviruses from a certain host order being scattered across the viral phylogeny (Figure 3). In addition, and in line with the viral sequence identity analysis presented above, the ambivirus phylogeny demonstrated that the majority of viruses discovered in this study constitute yet undescribed lineages. These undescribed lineages are distinct from known ambiviruses that are largely derived from metatranscriptomes (gray branches in Figure 3) and for which the host, therefore, remains unknown (Forgia et al., 2022; Lee et al., 2023). The discovery of 345 viral sequences from publicly available

unprocessed sequencing archives reinforced the notion that data-driven virus discovery approaches (Lauber and Seitz, 2022) open new opportunities for studying the natural diversity and evolution of viruses, viroids, and other infectious agents that exist on our planet at unprecedented detail and depth. The DDVD approach is uncoupled from the collection, processing, and sequencing of biological samples and, thus, allows for projects of a scale that conventional virus discovery studies cannot compete with.

To investigate the potential relevance of ambiviruses to human health and disease, we performed a second screen of 12,694 human metatranscriptome projects from the SRA. This data set included samples from various human body sites. Out of the more than 12,000 experiments screened, there were only three data sets with ambivirus sequences fulfilling our hit criteria during the virus identification stage of our workflow (at least five read pairs identified with an *E*-value of  $1e-5$  or lower). Two of them were from the same study analyzing lung



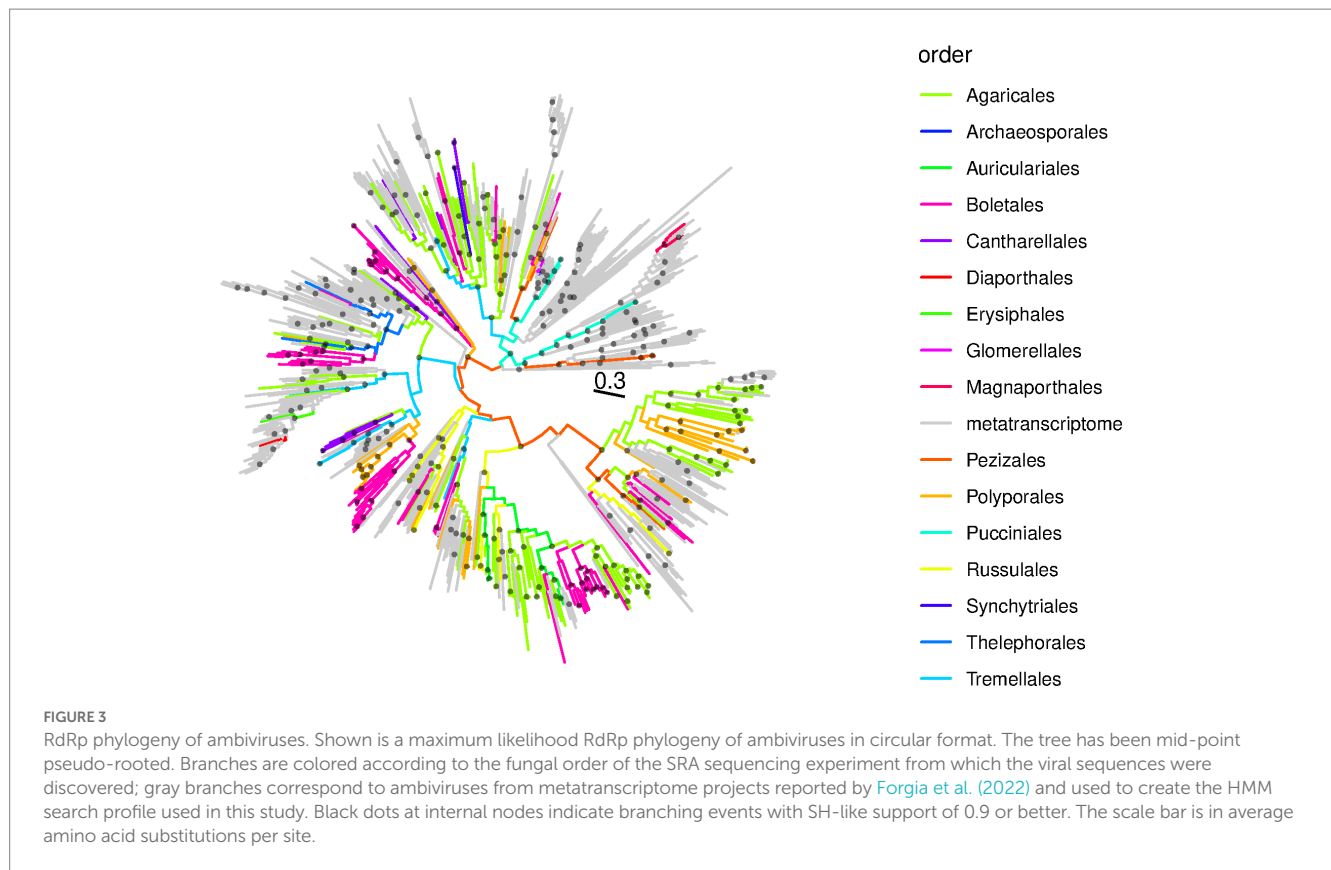
metatranscriptomes of patients with pneumonia and acute respiratory infections (SRA run accessions: SRR13677688 and SRR13677804). The third ambivirus-positive experiment (SRR5963935) was from a stool sample of a patient with Crohn's disease. The very low number of identified ambivirus makes it very challenging to discriminate between origin of the ambiviruses sequences by infection of fungi from the human microbiome and origin by any source of contamination (Cobbin et al., 2021). In general, the virtual absence of ambiviruses from human metatranscriptomes suggests that these viroid-like elements do not present a major factor interacting with the fungal part of the human microbiome (Pérez, 2021). We note that we cannot fully exclude the possibility that the observed lack of ambiviral sequences in the data is caused by an absence or strong underrepresentation of fungal hosts in the human microbiome samples. Indeed, an inspection of a selected set of human metatranscriptomic data sets using the Taxonomy Analysis Tool at the NCBI/SRA website showed that no reads were classified as fungal for some of them. However, each of the

SRA data sets inspected had a considerable fraction of “unidentified reads”, indicating that the reads cannot be assigned to any origin based on current reference sequences of known organisms. The proportion of unclassified reads varied greatly across data sets, and exceeded 98% in some cases (see for instance SRR935342). This observation reinforces the notion that our knowledge about the natural diversity of biological entities, both viral and cellular, remains incomplete.

## Conclusion

In summary, our study unveiled a large diversity of unknown ambivirus-like sequences in a large variety of fungi species. Future virus discovery efforts will show whether similar RdRp- and ribozyme-encoding hybrid elements exist in other hosts, including vertebrates and other animals. Studying the deep evolutionary relationships of this (and potentially other) distinct class(es) of





sub-viral elements with ancient and extant RNA viruses may offer unprecedented insights into the emergence of RNA viruses and their hallmark RdRp protein.

## Methods

### Sequence Read Archive data and metadata

A list of all 46,519 publicly available transcriptome experiments representing the full diversity of the kingdom *Fungi* (except those of the model organisms *Saccharomyces cerevisiae* and *Schizosaccharomyces pombe*) in the NCBI SRA database was compiled as of October 2022. The following search query was performed to obtain the SRA run identifiers: 'txid4751[Organism:exp] NOT txid4932[Organism:exp] NOT txid4896[Organism:exp] AND (cluster\_public[prop] AND "biomol rna"[Properties])'. A list of all 12,694 human metatranscriptome experiments were compiled from the SRA database using the following search query: 'txid1131769[Organism:exp] OR txid1504969[Organism:exp] OR txid1632839[Organism:exp] OR txid1633571[Organism:exp] OR txid1679718[Organism:exp] OR txid1712573[Organism:exp] OR txid1837932[Organism:exp] OR txid1842734[Organism:exp] OR txid2489051[Organism:exp] OR txid2705415[Organism:exp] OR txid408170[Organism:exp] OR txid433733[Organism:exp] OR txid447426[Organism:exp] OR txid539655[Organism:exp] OR txid646099[Organism:exp] AND (cluster\_public[prop] AND "biomol rna"[Properties])'. SRA data were downloaded using the SRA Toolkit ([Leinonen et al., 2011](#)). The taxonomic identifier for each SRA data set

was retrieved using the pysradb tool ([Choudhary, 2019](#)) while the taxonomic lineages information was fetched by the environment for tree exploration (ETE) toolkit ([Huerta-Cepas et al., 2016](#)). The host taxonomy information was reformatted into the Metaphlan2 format ([Truong et al., 2015](#)) using a customized script and the Pavian tool ([Breitwieser and Salzberg, 2020](#)) was used to produce Sankey diagrams.

### Sequence Read Archive-based virus discovery

The computational virus discovery workflow and its application to raw, unprocessed SRA data are described in previous studies ([Lauber et al., 2017, 2021](#)). The workflow is highly parallelized and was run on the high-performance computing cluster Taurus of the University of Technology (TU) Dresden. Here, we applied the Virushunter and Virusgatherer modules which screen a set of sequencing experiments from the SRA using one or several query pHMMs and perform a targeted, seed-based assembly of the identified sequencing experiments, respectively. The Virushunter performs a micro-assembly of sequencing read pairs identified in the pHMM search to create microcontigs that span the viral genome region covered by the query profile(s) (ambiviral RdRp in this study) or a part of that region. The Virusgatherer produces full-length or partial genome sequences depending on read coverage. Known ambiviruses previously discovered mostly in environmental metagenomes ([Forgia et al., 2022](#)) were used to construct a pHMM of ambivirus RdRp. We only considered hits for which at least five read pairs were identified with an E-value of  $1e-5$  or lower. Selected SRA data sets with high ambivirus read abundance were assembled *de novo*

using SPAdes in RNA mode (Bankevich et al., 2012) to independently validate the results of the seed-based assembly. Full-length circular RNAs were identified using vsearch (Lee et al., 2023).

The Virushunter and Virusgatherer tools as well as other code and further information are available on github: <https://github.com/lauberlab/VirusHunterGatherer> and [https://github.com/lauberlab/ambivirus\\_discovery\\_paper](https://github.com/lauberlab/ambivirus_discovery_paper).

## Open reading frame and RdRp identification

The presence of ORFs within the contig sequences was predicted using EMBOSS getorf (Rice et al., 2000). Only ORFs longer than 150 amino acids and inferred using the standard genetic code were considered for each circular RNA. Location of the RdRp was determined by comparing the *in silico* translated protein sequences encoded by the ORFs against the ambivirus RdRp profile with HMMER (Eddy, 2011).

## Ribozyme identification

The presence and genomic positions of ribozymes in ambivirus sequences were predicted using Infernal v1.1.4 (Nawrocki and Eddy, 2013) with the Rfam database (Kalvari et al., 2021). We considered hits with an E-value of 0.01 or lower.

## RNA secondary structure prediction

RNA secondary structure conformations of selected circular RNAs were predicted using RNAfold from the ViennaRNA package (Lorenz et al., 2011).

## Phylogenetic analysis

A multiple RdRp protein sequence alignment was computed using MAFFT v7.310 (Kato and Standley, 2013) with options ‘--localpair --maxiterate 1000’, followed by manual curation. We only kept well-conserved RdRp alignment positions with less than 50% of gaps across all sequences, allowing the inclusion of RdRp fragments that contributed many of the gaps. We used ModelTest-NG v0.1.7 (Darriba et al., 2020) to determine the best-fitting amino acid substitution model, which was LG+G4+I. Phylogenetic reconstruction was performed using PhyML v20120412 (Guindon et al., 2010). The tree was visualized using the ggtree R package (Yu et al., 2017).

A time-scaled phylogeny of fungal classes was obtained from TimeTree 5 (<http://timetree.org/>; Kumar et al., 2022).

## Statistical analysis

Contingency tables of the number of ambivirus-positive and -negative SRA experiments for a pair of host taxa to be compared and used Fisher’s exact test were compiled to assess the significance of differences. We used statistical functions (scipy.stats) in Python (Virtanen et al., 2020).

## Data availability statement

The original contributions presented in the study are included in the article/Supplementary material, further inquiries can be directed to the corresponding author.

## Author contributions

CL and LC performed experiments, analyzed the data and prepared figures. CL designed the study, supervised the project, and wrote the manuscript with contributions from LC. All authors contributed to the article and approved the submitted version.

## Funding

LC and CL are supported by the Project “Virological and immunological determinants of COVID-19 pathogenesis—lessons to get prepared for future pandemics (KA1-Co-02 ‘COVIPA’),” a grant from the Helmholtz Association’s Initiative and Networking Fund. CL was supported by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under Germany’s Excellence Strategy—EXC 2155—project number 390874280. This publication is funded by the Deutsche Forschungsgemeinschaft (DFG) as part of the “Open Access Publikationskosten” program.

## Acknowledgments

We thank all colleagues in the scientific community who make their sequencing data publicly accessible. We acknowledge the NCBI for providing an elaborate platform to exchange sequencing data. We thank the Center for Information Services and High-Performance Computing (ZIH) at TU Dresden for generous allocations of computer time. CL is a member of the European Virus Bioinformatics Center (EVBC).

## Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

## Publisher’s note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

## Supplementary material

The Supplementary material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fmicb.2023.1144003/full#supplementary-material>

## References

- Bankevich, A., Nurk, S., Antipov, D., Gurevich, A. A., Dvorkin, M., Kulikov, A. S., et al. (2012). SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. *J. Comput. Biol.* 19, 455–477. doi: 10.1089/cmb.2012.0021
- Branch, A. D., Benenfeld, B. J., and Robertson, H. D. (1988). Evidence for a single rolling circle in the replication of potato spindle tuber viroid. *Proc. Natl. Acad. Sci. U. S. A.* 85, 9128–9132. doi: 10.1073/pnas.85.23.9128
- Breitwieser, F. P., and Salzberg, S. L. (2020). Pavian: interactive analysis of metagenomics data for microbiome studies and pathogen identification. *Bioinformatics* 36, 1303–1304. doi: 10.1093/bioinformatics/btz715
- Bruening, G., Passmore, B. K., van Tol, H., Buzayan, J. M., and Feldstein, P. A. (1991). Replication of a plant virus satellite RNA: evidence favors transcription of circular templates of both polarities. *Mol. Plant-Microbe Interact.* 4, 219–225. doi: 10.1094/mpmi-4-219
- Bruenn, J. A. (2003). A structural and primary sequence comparison of the viral RNA-dependent RNA polymerases. *Nucleic Acids Res.* 31, 1821–1829. doi: 10.1093/nar/gkg277
- Choudhary, S. (2019). PySrad: a Python package to query next-generation sequencing metadata and data from NCBI sequence read archive. *F1000Research* 8:532. doi: 10.12688/f1000research.18676.1
- Cobbin, J. C., Charon, J., Harvey, E., Holmes, E. C., and Mahar, J. E. (2021). Current challenges to virus discovery by meta-transcriptomics. *Curr. Opin. Virol.* 51, 48–55. doi: 10.1016/j.coviro.2021.09.007
- Daròs, J., Elena, S. F., and Flores, R. (2006). Viroids: an Ariadne's thread into the RNA labyrinth. *EMBO Rep.* 7, 593–598. doi: 10.1038/sj.embor.7400706
- Daròs, J. A., and Flores, R. (1995). Identification of a retroviroid-like element from plants. *Proc. Natl. Acad. Sci.* 92, 6856–6860. doi: 10.1073/pnas.92.15.6856
- Darriba, D., Posada, D., Kozlov, A. M., Stamatakis, A., Morel, B., and Flouri, T. (2020). ModelTest-NG: a new and scalable tool for the selection of DNA and protein evolutionary models. *Mol. Biol. Evol.* 37, 291–294. doi: 10.1093/molbev/msz189
- de la Peña, M., Ceprián, R., and Cervera, A. (2020). A singular and widespread group of mobile genetic elements: RNA circles with autocatalytic ribozymes. *Cells* 9:2555. doi: 10.3390/cells9122555
- Diener, T. O. (2001). The viroid: biological oddity or evolutionary fossil? *Adv. Virus Res.* 57, 137–184. doi: 10.1016/s0065-3527(01)57003-7
- Eddy, S. R. (2011). Accelerated profile HMM searches. *PLoS Comput. Biol.* 7:e1002195. doi: 10.1371/journal.pcbi.1002195
- Ferre-D'Amare, A. R., and Scott, W. G. (2010). Small self-cleaving ribozymes. *Cold Spring Harb. Perspect. Biol.* 2:a003574. doi: 10.1101/cshperspect.a003574
- Forgia, M., Isgandarli, E., Aghayeva, D. N., Huseynova, I., and Turina, M. (2021). Virome characterization of *Cryphonectria parasitica* isolates from Azerbaijan unveiled a new myomonavirus and a putative new RNA virus unrelated to described viral sequences. *Virology* 553, 51–61. doi: 10.1016/j.virol.2020.10.008
- Forgia, M., Navarro, B., Daghino, S., Cervera, A., Gisel, A., Perotto, S., et al. (2022). Extant hybrids of RNA viruses and viroid-like elements. *bioRxiv*. doi: 10.1101/2022.08.21.504695
- Giguère, T., Raj Adkar-Purushothama, C., and Perreault, J.-P. (2014). Comprehensive secondary structure elucidation of four genera of the family Pospiviroidae. *PLoS One* 9:e98655. doi: 10.1371/journal.pone.0098655
- Gorbalenya, A. E., Pringle, F. M., Zeddam, J.-L., Luke, B. T., Cameron, C. E., Kalkmakoff, J., et al. (2002). The palm subdomain-based active site is internally permuted in viral RNA-dependent RNA polymerases of an ancient lineage. *J. Mol. Biol.* 324, 47–62. doi: 10.1016/s0022-2836(02)01033-1
- Guindon, S., Dufayard, J.-F., Lefort, V., Anisimova, M., Hordijk, W., and Gascuel, O. (2010). New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. *Syst. Biol.* 59, 307–321. doi: 10.1093/sysbio/syq010
- Huerta-Cepas, J., Serra, F., and Bork, P. (2016). ETE 3: reconstruction, analysis, and visualization of Phylogenomic data. *Mol. Biol. Evol.* 33, 1635–1638. doi: 10.1093/molbev/msw046
- Kalvari, I., Nawrocki, E. P., Ontiveros-Palacios, N., Argasinska, J., Lamkiewicz, K., Marz, M., et al. (2021). Rfam 14: expanded coverage of metagenomic, viral and micro RNA families. *Nucleic Acids Res.* 49, D192–D200. doi: 10.1093/nar/gkaa1047
- Katoh, K., and Standley, D. M. (2013). MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol. Biol. Evol.* 30, 772–780. doi: 10.1093/molbev/mst010
- Kos, A., Dijkema, R., Arnberg, A. C., van der Meide, P. H., and Schellekens, H. (1986). The hepatitis delta (delta) virus possesses a circular RNA. *Nature* 323, 558–560. doi: 10.1038/323558a0
- Kumar, S., Suleski, M., Craig, J. M., Kasprovicz, A. E., Sanderford, M., Li, M., et al. (2022). TimeTree 5: an expanded resource for species divergence times. *Mol. Biol. Evol.* 39:msac174. doi: 10.1093/molbev/msac174
- Lauber, C., and Seitz, S. (2022). Opportunities and challenges of data-driven virus discovery. *Biomolecules* 12:1073. doi: 10.3390/biom12081073
- Lauber, C., Seitz, S., Mattei, S., Suh, A., Beck, J., Herstein, J., et al. (2017). Deciphering the origin and evolution of hepatitis B viruses by means of a family of non-enveloped fish viruses. *Cell Host Microbe* 22, 387–399.e6. doi: 10.1016/j.chom.2017.07.019
- Lauber, C., Vaas, J., Klingler, F., Mutz, P., Gorbalenya, A. E., Bartenschlager, R., et al. (2021). Deep mining of the sequence read archive reveals bipartite coronavirus genomes and inter-family spike glycoprotein recombination. *bioRxiv*. doi: 10.1101/2021.10.20.465146
- Lee, B. D., Neri, U., Roux, S., Wolf, Y. I., Camargo, A. P., Krupovic, M., et al. (2023). Mining metatranscriptomes reveals a vast world of viroid-like circular RNAs. *Cells* 186, 646–661.e4. doi: 10.1016/j.cell.2022.12.039
- Leinonen, R., Sugawara, H., and Shumway, M. International Nucleotide Sequence Database Collaboration (2011). The sequence read archive. *Nucleic Acids Res.* 39, D19–D21. doi: 10.1093/nar/gkq1019
- Lorenz, R., Bernhart, S. H., Höner Zu Siederdisen, C., Tafer, H., Flamm, C., Stadler, P. F., et al. (2011). ViennaRNA Package 2.0. *Algorithms Mol. Biol.* 6:26. doi: 10.1186/1748-7188-6-26
- Modahl, L. E., Macnaughton, T. B., Zhu, N., Johnson, D. L., and Lai, M. M. (2000). RNA-dependent replication and transcription of hepatitis delta virus RNA involve distinct cellular RNA polymerases. *Mol. Cell. Biol.* 20, 6030–6039. doi: 10.1128/MCB.20.16.6030-6039.2000
- Naranjo-Ortiz, M. A., and Gabaldón, T. (2019). Fungal evolution: diversity, taxonomy and phylogeny of the Fungi. *Biol. Rev.* 94, 2101–2137. doi: 10.1111/brv.12550
- Nawrocki, E. P., and Eddy, S. R. (2013). Infernal 1.1: 100-fold faster RNA homology searches. *Bioinformatics* 29, 2933–2935. doi: 10.1093/bioinformatics/btt509
- Nguyen, L. D. N., Viscogliosi, E., and Delhaes, L. (2015). The lung mycobiome: an emerging field of the human respiratory microbiome. *Front. Microbiol.* 6:89. doi: 10.3389/fmicb.2015.00089
- Pérez, J. C. (2021). Fungi of the human gut microbiota: roles and significance. *Int. J. Med. Microbiol.* 311:151490. doi: 10.1016/j.ijmm.2021.151490
- Rao, A. L. N., and Kalantidis, K. (2015). Virus-associated small satellite RNAs and viroids display similarities in their replication strategies. *Virology* 479–480, 627–636. doi: 10.1016/j.virol.2015.02.018
- Rice, P., Longden, I., and Bleasby, A. (2000). EMBOS: the European molecular biology open software suite. *Trends Genet.* 16, 276–277. doi: 10.1016/S0168-9525(00)02024-2
- Soret, P., Vandenborgh, L.-E., Francis, F., Coron, N., Enaud, R., Avalos, M., et al. (2020). Respiratory mycobiome and suggestion of inter-kingdom network during acute pulmonary exacerbation in cystic fibrosis. *Sci. Rep.* 10:3589. doi: 10.1038/s41598-020-60015-4
- Sureau, C., and Negro, F. (2016). The hepatitis delta virus: replication and pathogenesis. *J. Hepatol.* 64, S102–S116. doi: 10.1016/j.jhep.2016.02.013
- Sutela, S., Forgia, M., Vainio, E. J., Chiappello, M., Daghino, S., Vallino, M., et al. (2020). The virome from a collection of endomycorrhizal fungi reveals new viral taxa with unprecedented genome organization. *Virus Evolution* 6:veaa076. doi: 10.1093/ve/veaa076
- Truong, D. T., Franzosa, E. A., Tickle, T. L., Scholz, M., Weingart, G., Pasolli, E., et al. (2015). MetaPhlan2 for enhanced metagenomic taxonomic profiling. *Nat. Methods* 12, 902–903. doi: 10.1038/nmeth.3589
- Virtanen, P., Gommers, R., Oliphant, T. E., Haberland, M., Reddy, T., Cournapeau, D., et al. (2020). SciPy 1.0: fundamental algorithms for scientific computing in Python. *Nat. Methods* 17, 261–272. doi: 10.1038/s41592-019-0686-2
- Wang, Y. (2021). Current view and perspectives in viroid replication. *Curr. Opin. Virol.* 47, 32–37. doi: 10.1016/j.coviro.2020.12.004
- Yu, G., Smith, D. K., Zhu, H., Guan, Y., and Lam, T. T. (2017). Ggtree: an R package for visualization and annotation of phylogenetic trees with their covariates and other associated data. *Methods Ecol. Evol.* 8, 28–36. doi: 10.1111/2041-210X.12628
- Zhang, F., Aschenbrenner, D., Yoo, J. Y., and Zuo, T. (2022). The gut mycobiome in health, disease, and clinical applications in association with the gut bacterial microbiome assembly. *The Lancet Microbe* 3, e969–e983. doi: 10.1016/S2666-5247(22)00203-8