

RESEARCH

Open Access



# Genomic insights reveal community structure and phylogenetic associations of endohyphal bacteria and viruses in fungal endophytes

Efraín Escudero-Leyva<sup>1,2</sup>, Michal Belle<sup>3</sup>, Abolfazl DadkhahTehrani<sup>3</sup>, James N. Culver<sup>4</sup>, Marcelo Araya-Salas<sup>5,6</sup>, Joseph P. Kutza<sup>4</sup>, Natasha Goldson<sup>3</sup>, Max Chavarría<sup>1,2,7</sup> and Priscila Chaverri<sup>3\*</sup>

## Abstract

**Background** Endohyphal microbial communities, composed of bacteria and viruses residing within fungal hyphae, play important roles in shaping fungal phenotypes, host interactions, and ecological functions. While endohyphal bacteria have been shown to influence fungal pathogenicity, secondary metabolism, and adaptability, much remains unknown about their diversity and host specificity. Even less is known about endohyphal viruses, whose ecological roles and evolutionary dynamics are poorly understood. This study integrates genomic and transcriptomic approaches to (1) characterize the diversity of endohyphal bacterial and viral communities in fungal endophytes isolated from *Fagus grandifolia* leaves, and (2) assess potential host specialization through phylogenetic signal analyses.

**Results** We analyzed 19 fungal isolates spanning eight fungal orders (*Amphisphaeriales*, *Botryosphaeriales*, *Diaporthales*, *Glomerellales*, *Mucorales*, *Pleosporales*, *Sordariales*, and *Xylariales*). Bacterial communities were highly diverse and showed significant phylogenetic signal, with core taxa—such as *Bacillales*, *Burkholderiales*, *Enterobacterales*, *Hyphomicrobiales*, and *Pseudomonadales*—shared across hosts. Several bacterial groups were associated with specific fungal orders, suggesting host specialization: *Moraxellales*, *Sphingomonadales*, and *Streptosporangiaceae* in *Amphisphaeriales*; *Enterobacterales*, *Hyphomicrobiales*, and *Micrococcales* in *Glomerellales*; and *Cytophagales* in *Diaporthales*. In contrast, viral communities were less diverse and dominated by double-stranded DNA viruses, primarily *Bamfordvirae* and *Heunggongvirae*. No core viral taxa were detected in metatranscriptomic data, and only a few reads of double-stranded RNA viruses were found.

**Conclusions** Overall, our results indicate potential host specialization in bacterial endophytes and limited viral diversity in fungal hosts, with dsDNA viruses dominating the endohyphal virome. These findings provide new insights into the ecological and evolutionary dynamics of fungal-associated microbiota. Future work expanding taxonomic reference databases and exploring the functional roles of these microbial symbionts will be essential to understanding their contributions to fungal biology, host interactions, and broader ecosystem processes.

**Keywords** Bacteriome, Fungi, Holobiont, Mycobiome, Mycoviruses, Symbiosis, Virome

\*Correspondence:  
Priscila Chaverri  
pchaverri@bowiestate.edu

Full list of author information is available at the end of the article



© The Author(s) 2025. **Open Access** This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

## Background

Symbiotic relationships within an ecological network can be classified as direct or indirect [1, 2]. A direct relationship involves the immediate effect of one organism on another when not mediated or transmitted through a third individual. An example of a direct interaction is fungal endophytism, where fungi live within healthy plant tissues without causing disease symptoms. Some endophytes may eventually become pathogens [3], while others remain commensals or mutualists, aiding plant growth and offering protection against diseases or abiotic stressors [4]. In contrast, an indirect relationship can be exemplified by the endohyphal microbiota—microorganisms residing within the hyphae of endophytic fungi. These microorganisms can have positive or negative effects on their fungal host, which in turn, indirectly influences plant biology [5–7]. Disentangling the plant's holobiont and its symbiotic relationships is crucial for understanding how the presence or absence of one or multiple symbionts can trigger cascading effects, ultimately influencing the fate of plants and, hence, plant communities and ecosystems [8, 9].

Despite advancements in tools to more comprehensively characterize microbial communities (e.g., metagenomics, metatranscriptomics, and bioinformatics), the full spectrum of the multi-species and multi-level symbiotic relationships within plant-associated fungi (e.g., endophytes) remains poorly known. A few studies have revealed the complexity of these endohyphal communities, which include bacteria, viruses, and, occasionally, microalgae [6, 10, 11]. However, basic knowledge, such as alpha or beta diversity within fungal hyphae and across fungal taxa, especially for endohyphal viruses, is limited.

Endohyphal bacteria have been studied more extensively than viruses (e.g. [12–16]). Many studies have focused on the effects they can induce in their fungal hosts [5, 17]. For instance, they can reduce or increase pathogenicity [18, 19], increase spore and mycotoxin production [20], heighten respiration and hyphal density [21], protect against predatory nematodes and amoebae [22], and support endophytic establishment [23, 24], among others. Other studies have focused on characterizing their diversity. For example, a large-scale 16S ribosomal DNA metabarcoding study that screened hundreds of fungal isolates across several taxonomic levels revealed bacterial associations in nearly 90% of the isolates [11], in contrast to the ~20% found by Hoffman & Arnold [13]. However, most published studies have concentrated on a few non-*Dikarya* fungal groups (e.g., *Glomeromycota*, *Mortierellomycota*, and *Mucoromycota*), with even less attention given to other highly diverse fungal *Dikarya* phyla (e.g., *Ascomycota* and *Basidiomycota*) [6], highlighting the opportunity to explore and uncover a vast array of uncharacterized diversity.

The diversity and function of endohyphal viruses remain largely unexplored compared to endohyphal bacteria. Although most research has focused on mycoviruses in pathogenic *Ascomycota* and *Basidiomycota*, these viruses are present across all fungal lineages [25]. Recent advancements in high-throughput sequencing technologies have revealed that the fungal virome encompasses a variety of genome types, altering the previous misconception that mycoviruses were primarily double-stranded RNA (dsRNA) [26–29]. Mycoviral diversity includes more than 25 families, many of which have RNA genomes (e.g., dsRNA), and others with DNA genomes. The functions of most mycoviruses are poorly known, except for some dsRNA viruses. For example, several dsRNA viruses can confer hypovirulence, making them potential candidates for the biocontrol of mycotoxigenic or plant-pathogenic fungi [29–34]. These dsRNA hypovirulence mycoviruses can also transform non-pathogenic endophytic fungi (e.g., *Colletotrichum*, *Pestalotiopsis*, and *Sclerotinia*, among others) into highly phytopathogenic fungi [34–38]. Conversely, hypervirulence has emerged as an attractive strategy to enhance the biocontrol effectiveness of entomopathogenic fungi, such as *Beauveria bassiana* and its associated virus, '*Beauveria bassiana* victorivirus 1 (BbVV-1)' [32, 39].

Bacteria can associate with fungi both externally and internally, via horizontal or vertical transmission, with the nature of the interaction influenced by the fungal morphology as well as the surface molecules and secreted factors of both the fungi and bacteria [40]. Therefore, one would expect that host-specificity, limited host ranges, or coevolution would be more widespread in fungi-bacteria interactions, fostering increased species diversity by promoting niche specialization and enabling resource partitioning, among other factors [41–43]. However, studies have reported that while certain bacterial groups, such as *Pseudomonadota* (= *Proteobacteria*), *Actinomycetota* (= *Actinobacteria*), and *Bacillota*, tend to associate with most fungi, host specificity is not always evident [11, 13, 14]. Some well-known examples of obligate symbiosis include *Burkholderiales* (*Betaproteobacteria*, *Pseudomonadota*) with *Mucoromycota* (e.g., fungus *Rhizopus* - bacterium *Mycetohabitans*), *Glomeromycota* (e.g., *Gigaspora* - '*Candidatus Glomeribacter gigasporum*'), and *Mortierellomycota* (e.g., *Mortierella* - *Mycoavidus*) [40].

Mycoviruses may be transmitted through fungal spores (vertical transmission) or hyphal anastomosis (horizontal transmission) [10, 44], with no known extracellular entry route [45]. This lack of an extracellular entry route has been a significant challenge for the intended propagation of mycoviruses and their effective use in biological control [46]. Additionally, some viruses (e.g., dsDNA viruses) have incorporated viral elements into fungal genomes through mechanisms that are not yet fully understood,

occurring either in recent or ancient (millions of years) events [47–49]. The above-mentioned transmission mechanisms suggest the potential for niche specialization and/or coevolution, which could then result in some level of host-specificity [41–43]. However, no host-specificity or preference has been reported to date.

Based on what is known of bacterial and viral transmission in fungi, we hypothesize microbial communities to be more similar between closely related fungal hosts as a result of host specialization. This pattern would reflect a phylogenetic signal, where the phylogenetic relatedness of fungal hosts is associated with the composition of their endohyphal bacterial and viral communities. Therefore, the aims of this study were two-fold. First, we aimed to characterize the bacterial and viral genomic and transcriptomic diversity residing in the hyphae of selected endophytes isolated from *Fagus grandifolia* (American beech) using *de novo* metagenomics and metatranscriptomics. Second, with the data obtained from the first objective, we evaluated the phylogenetic signal of microbial communities and core microbial taxa. The results of this study will enhance our understanding of microbial diversity, particularly how it may influence the phenotype and genotype of the host fungus and its indirect interactions with plants. This knowledge is crucial for elucidating the complex multi-species and multi-level symbiotic relationships within plant-associated fungi and their potential applications in pathogen-specific biocontrol and ecosystem health.

## Methods

### Fungal isolation and identification

Endophytic fungi were obtained following previously used protocols [50, 51], which are described below. These were isolated from two *Fagus grandifolia* trees located in the forest adjacent to Bowie State University (39° 01' 17.69" N; -76° 45' 24.62" W), which has a humid subtropical climate [52]; five healthy leaves per tree were collected. Five discs (5 mm in diameter) per leaf were excised and surface-sterilized by sequential immersion in bleach (2% for 30 s), ethanol (70% for 1 min), and thoroughly rinsed with sterile distilled water. The sterilized discs were placed in Petri dishes (10 cm) containing potato dextrose agar (PDA, 15 mL, Difco, Detroit, MI, USA). To inhibit extracellular and free-living endophytic bacteria while minimizing effects on endohyphal bacteria, a low concentration (1%) and combination of antibiotics (i.e., neomycin-penicillin-streptomycin; Sigma-Aldrich, St. Louis, MO, USA) that do not easily permeate into the fungal cell wall was added to the PDA only in the initial fungal isolation from leaf tissues. Plates were incubated (25 °C). Subculturing was performed immediately once hyphal growth began, transferring the mycelium to fresh PDA plates without antibiotics to

establish axenic cultures. Surface sterilization effectiveness was verified using the leaf imprint method [53], as previously described and applied in our earlier work [51]. Fungal cultures are currently stored (-80 °C) in cryotubes (2 mL) with glycerol (20%).

DNA extraction was done after approximately 7 days of growth on PDA. The mycelium was harvested and processed using the PowerPlant Mini kit (Qiagen Inc., Hilden, Germany) following the manufacturer's instructions. The internal transcribed spacers region (ITS1, 5.8 S, ITS2) of the nuclear ribosomal DNA (ITS nrDNA) was amplified using primers ITS4 and ITS5 [54]. Additional gene regions were amplified to refine the identification for selected isolates: translation-elongation factor 1- $\alpha$  (TEF) using primers EF728f and EF2r [55], and  $\beta$ -tubulin (TUB) using primers Bt1 and Bt2 [56]. The PCR reaction mixture consisted of GoTaq Green Master Mix (12.5  $\mu$ L; Promega Corporation, Madison, WI, USA), forward and reverse primers (1  $\mu$ L of each), bovine serum albumin (BSA, 0.5  $\mu$ L, ThermoFisher Scientific, Waltham, MA, USA), dimethyl sulfoxide (DMSO, 1  $\mu$ L, Merck, Darmstadt, Germany), UltraPure nuclease-free water (7  $\mu$ L, ThermoFisher Scientific) and the template DNA (2  $\mu$ L) [57]. PCR products were purified and sequenced by Psomagen (Maryland, USA). The sequences were trimmed and aligned using BioEdit v.7.7.1 [58], and BLASTn searches were conducted against the NCBI GenBank database for each gene region. Fungal identification was based on sequences with >98% identity and >85% coverage [59], prioritizing the most complete taxonomic classification available. Newly generated sequences were deposited in GenBank (Supplementary Table S1).

Once all collected endophyte isolates were identified, we selected samples that represented the diversity found in *F. grandifolia* leaves, ensuring at least four representatives of some fungal orders and multiple isolates from the same fungal family to test for host association. After this selection, 19 isolates remained for the subsequent analyses: *Amphisphaeriales* (4 isolates), *Diaporthales* (5), *Glomerellales* (4), which were those with more representatives; and then *Botryosphaeriales* (1), *Mucorales* (1), *Pleosporales* (1), *Sordariales* (1), and *Xylariales* (2), which had fewer isolates (Supplementary Table S1). Of those, the best-represented families were *Diaporthaceae*, *Glomerellaceae*, and *Pestalotiopsidaceae*; *Diaporthe*, *Colletotrichum*, and *Pestalotiopsis* as their corresponding genera.

### Fungal phylogenetic analysis

To be able to test for phylogenetic signals, we first had to reconstruct a phylogeny for the fungal isolates selected. For the selected 19 samples, sequences from each gene were independently aligned using MUSCLE within MEGA v.11 [60, 61]. To ensure comparable sequence

lengths, the alignments were trimmed: ITS nrDNA alignment to 676 bp, TEF to 398 bp, and TUB to 404 bp. Then, the trimmed alignments were concatenated using MEGA and exported to Phylip format for phylogenetic tree reconstruction. Prior to building the tree, ModelFinder [62] was used inside IQ-TREE v.2.2.0 [63] to identify the most suitable substitution model for the concatenated dataset. To assess the robustness of the inferred relationships, 1000 bootstrap replicates were performed. This analysis was done using the Kabré HPC cluster (CeNAT-CONARE, Costa Rica). Finally, the consensus tree was visualized using FigTree v.1.4.4 (<http://tree.bio.ed.ac.uk/software/figtree/>) setting sample M22 (*Umbelopsis* aff. *isabellina*, *Mucoromycota*) as the outgroup.

#### Total DNA and RNA extraction and sequencing

Following purification, fungal isolates were grown in liquid culture. Erlenmeyer flasks (250 mL) containing malt extract broth (MEB, 50 mL, Difco, Detroit, MI, USA) were inoculated with five mycelial plugs (5 mm diameter) and placed in an I24 New Brunswick Scientific incubator shaker (New Brunswick Scientific, New Jersey, USA) (25 °C, 110 rpm) for seven days. Before harvesting the mycelium, we confirmed that the broth was clear and free of bacterial contaminants. Fresh mycelium was harvested using a Buchner funnel system, sterile filter paper, and vacuum, then dried with sterile paper towels, immediately flash-frozen with liquid nitrogen, and stored at -80 °C for a maximum of five days. Approximately 100 mg of frozen mycelia were then ground using a mortar and pestle with liquid nitrogen throughout the process. Each sample was ground separately at different times to avoid cross-contamination. Then, total DNA and RNA were extracted from the ground mycelia using commercially available kits (DNeasy Plant Pro Kit and RNeasy Plant Mini Kit; Qiagen Inc., Hilden, Germany) following the manufacturer's instructions. Total RNA and DNA were further cleaned and concentrated using DNA and RNA Clean and Concentration kits (Zymo Research Corporation, Irvine, California, USA) according to the manufacturer's protocol. The purified DNA and RNA samples were then stored (-80 °C). Total DNA and RNA were sent to Novogene Inc. (California, USA) for shotgun metagenomic and metatranscriptomic sequencing (NovaSeq PE150; 6Gb per sample; ribosomal depletion for RNA samples). All raw data (.fastq files) have been deposited in GenBank under BioProject PRJNA1221291.

#### Bioinformatic analyses of metagenomic and metatranscriptomic data

Computational analyses were performed on the Kabré HPC Cluster. Data quality control and filtering were done by Novogene, including adapter and low-quality sequence removal, resulting in raw data quality >92%

Q30. Reads were further filtered and trimmed with Seqtk v.1.4 (<https://github.com/lh3/seqtk/>) and BBduk v.38.84 (<https://sourceforge.net/projects/bbmap/>). To attempt to capture low-abundance taxa and enable more accurate taxonomic classification, further analyses were run and compared using both reads (unassembled) and contigs (assembled). Assembly of the metagenomic and metatranscriptomic data was done using metaSPAdes [64] and rnaSPAdes [65], respectively. After assembly and before running the taxonomy classifier on metagenome and metatranscriptome assemblies, contigs were filtered to a minimum length of 500 bp. Taxonomic assignments were made with Kaiju [66], selecting the “nr” v.2023-05-10 and “virus” v.2023-05-10 databases for prokaryotes and viruses, respectively, using the default parameters (“greedy”; allowed mismatches = 3; e-value = 0.01; minimum match length = 11–15; minimum match score = 65–75). Kaiju is considered a sensitive taxonomy classifier that uses NCBI RefSeq database and protein-based classifiers [66–68]. Kaiju was run for prokaryotic and viral reads and contigs with metagenomic data, and only for viral reads and contigs for metatranscriptomic data. The latter was done to determine if this approach would better detect dsRNA viruses. Counts and taxonomy tables generated by Kaiju were manually inspected and curated to remove potential artifacts and ensure data quality.

#### Transmission electron microscopy

Transmission electron microscopy (TEM) was performed with selected fungal isolates as another tool to investigate the presence of potential endohyphal bacteria and viruses. Isolates K21, M5t, and M67 were chosen for their apparent high number of anticipated viral contigs and taxa when compared to other isolates. These three isolates were cultured in MEB and shaken (~110 rpm, 20 °C) for 7 days in the dark. A conventional fixation and embedding method was used as described previously [69] with the following specifications. Hyphae from each culture were excised in fixative (4% paraformaldehyde and 3% glutaraldehyde in sodium phosphate buffer 0.1 mol/L, pH 7.1), washed three times in fixative, and stored (4 °C) for two days in the same fixative. Post-fixation was performed (1% osmium tetroxide for 2 h) and then washed with distilled water (10–15 min). Dehydration of samples was performed by a series of 15-minute ethanol treatments at concentrations of 25%, 50%, 75%, and 90%, and then in absolute ethanol (1 h). Dehydrated samples were infiltrated with absolute ethanol: Spurr resin [70] in mixtures (3:1, 1:1, 1:3 for 1 h each), then in pure resin (12 h). The polymerization of the resin was achieved at 60 °C for 24 h. Fully polymerized samples were ultrathin sectioned (60–100 nm) and stained with aqueous uranyl acetate (1–2%) and lead citrate (3%) [71]. Thin sections were



viewed in an HT7700 Transmission Electron Microscope (Hitachi, Japan; acceleration voltage up to 80 kV).

### Community analyses

To evaluate the composition of the endohyphal microbiota, metagenomic and metatranscriptomic contig and read data were analyzed using R v.4.3.3 [72]. First, to assess the extent of the sampling effort, taxa accumulation curves were generated with the iNEXT package using sequence abundance [73]. For beta diversity, two analyses were done: (1) with all the samples and (2) with samples with major representation to create a balanced dataset. This second dataset included isolates from *Amphisphaeriales* (K2, K5, K10, K18), *Diaporthales* (M1b, M13t, M20, M24, M32) and *Glomerellales* (M19, M27, M48, M66). For both datasets, the count data were transformed using the Hellinger method with the “decostand” function from the vegan package v.2.6.4 [74]. To evaluate differences among fungal orders and families, as well as differences across isolates, an ANOVA was performed for the full dataset and PERMANOVA for the balanced dataset, and a multivariate test was done for both using “adonis” and “betadisper” functions from vegan. Subsequently, non-metric multidimensional scaling (NMDS) was performed using a Bray-Curtis transformed matrix with vegan using “metaMDS” function. Additionally, indicator species analysis was performed using indicspecies package v.1.7.15 through the function “multipat” (multilevel pattern analysis) to retrieve possible bacterial or viral taxa important to the fungal orders using the species occurrence variable [75]. Then, to capture the overall structure of the community [76], a multinomial species classification was run with the function “clamtest” from the vegan package. Finally, to determine the core microbiome, taxa that were found in > 80% of the samples and with a relative abundance greater than 0.1% were considered “core,” and those present in 50–79% of the samples were considered “resident” [77–80]. Relative abundance (%) per sample was calculated and graphed in the R package microeco [81].

### Phylogenetic signal analyses

To assess a potential phylogenetic signal between endohyphal communities and their fungal hosts, a Mantel correlation test between phylogenetic distance and community similarity was conducted for the full and balanced datasets in R v.4.3.3 using the vegan package [72, 74]. The maximum likelihood matrix derived from the concatenated phylogenetic tree generated with IQ-TREE, alongside abundance tables for each community (bacteria and viruses) and technique (metagenomic and metatranscriptomic). The abundance tables were converted to presence/absence matrices, followed by a Jaccard transformation using the “vegdist” function from vegan. The

Mantel test was then run with 10,000 permutations. To validate the model, simulations were performed using the MASS package [82] randomly assigning different combinations of endohyphal communities and fungal hosts across 1,000 replicates. This approach was used to evaluate Type I and Type II errors. The script for these analyses can be consulted in the following link (<https://rpubs.com/marcelo-araya-salas/1218708>). Additionally, the Procrustean superimposition approach (Procrustes) [83] was performed using the NMDS values and the distance-based redundancy analysis (dbRDA) to evaluate the association of the communities within the order taxonomical level of fungi. This was done first by calculating the dbRDA values with the function “dbRDA” from the vegan package and then using the function “protest” in vegan.

## Results

### Fungal isolate identification and phylogeny

From the leaves of two *F. grandifolia* trees, 97 endophytic fungal isolates were recovered: 55 *Sordariomycetes* (*Apiospora*, *Beltrania*, *Colletotrichum*, *Diaporthe*, *Neopestalotiopsis*, *Nigrospora*, *Pestalotiopsis*, *Tubakia*, and unidentified *Xylariales*), 41 *Dothideomycetes* (*Cladosporium*, *Didymosphaeria*, and *Lasiodiplodia*), and one *Mucoromycetes* (*Umbelopsis*). (results not shown). Out of those, 19 isolates (already described in the Methods) were selected; most of them *Sordariomycetes*: 4 *Amphisphaeriales*, 5 *Diaporthales*, 4 *Glomerellales*, 1 *Botryosphaeriales*, 1 *Mucorales*, 1 *Pleosporales*, 1 *Sordariales*, and 2 *Xylariales* (Supplementary Table S1). The concatenated phylogenetic tree formed mostly well-supported clades that, in general, corresponded to the fungal families and orders (Supplementary Figure S1). All the supplementary figures and tables are publicly available at <https://doi.org/10.5281/zenodo.15620648>.

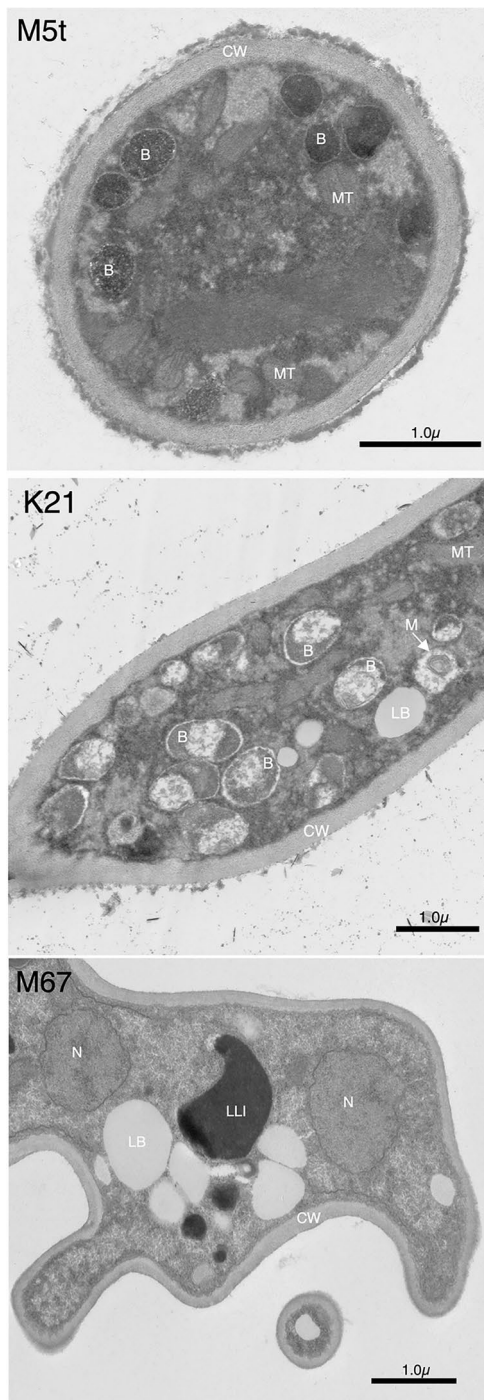
### Transmission electron microscopy

TEM analysis identified multiple bacteria-like cells within M5t and K21 isolates but not within M67 (Fig. 1). TEM images of the three fungal isolates did not identify clear virus-like particles.

### Bacterial metagenomic data analyses

#### Alpha diversity

A total of 15,979,400 reads matched prokaryotes using the Kaiju taxonomy classifier (Supplementary Table S2). Most samples yielded over 2,000 taxa, with M5t showing the lowest count at 1,962 (Supplementary Table S3). At the order level, approximately 3,000 taxa were identified for *Amphisphaeriales*, *Diaporthales*, and *Glomerellales*, while other orders registered around 2,000 taxa (Supplementary Table S3). Based on extrapolated data, *Amphisphaeriales*, *Diaporthales*, and *Glomerellales* appeared



**Fig. 1** Transmission Electron Microscopy (TEM) image of hyphal cross-sections of isolates M5t, K21 and M67. *B*: Endohyphal bacteria; *CW*: Fungal cell wall; *LB*: Fungal lipidic body; *LLI*: Fungal Lipid-like inclusion; *M*: Bacterial mesosome; *MT*: Fungal mitochondria. Fully polymerized samples were ultrathin sectioned (60–100 nm) and stained with aqueous uranyl acetate (1–2%) and lead citrate (3%)

to approach the asymptote (Supplementary Figure S2). At the family level, *Diaporthaceae* had the highest taxonomic count with approximately 3,000 taxa, followed by *Pestalotiopsidaceae* with nearly 2,900; the remaining

families had around 2,000 taxa (Supplementary Table S3). Among all families, only *Apiosporaceae* approached the asymptote based on both inter- and extrapolated data (Supplementary Figure S2).

After assembly and taxonomic classification, 7,747 metagenomic contigs were matched to bacteria using the Kaiju taxonomy classifier (Supplementary Table S4). Variability in contig counts was observed across samples, with most samples generating approximately 2,000 contigs and representing nearly 400 taxa. An exception was sample M22, which had significantly fewer contigs (Supplementary Figure S3A). When classified by fungal order, *Xylariales* was particularly notable, representing over 500 bacterial taxa across 2,000 contigs, despite being derived from only two samples. Most other orders, except *Botryosphaeriales* and *Mucorales*, comprised fewer than 300 taxa. *Botryosphaeriales* and *Mucorales* accounted for fewer than 250 taxa across 1,000 contigs (Supplementary Figure S3B). At the family level, most groups contained approximately 400 taxa, with *Umbelopsidaceae* showing the lowest number (<200). Extrapolated data suggested that *Apiosporaceae* might contain nearly 3,000 contigs, potentially representing up to 600 taxa. For other families, the data followed a similar trend, with 2,000 contigs yielding around 300 taxa (Supplementary Figure S3C). However, none of the isolates across the taxonomical classifications reached an asymptote, suggesting that bacterial richness might be even greater than observed.

#### **Bacterial community structure, composition, and phylogenetic signal across fungal hosts (metagenomics)**

The results of community structure and composition analyses yielded different results depending on whether reads or contigs were used. A summary of these findings, including p-values and other statistics, is provided in Table 1. When contigs were analyzed across all fungal orders, significant differences in microbial community composition were observed, along with heterogeneity among groups. Similar patterns were observed at the family level, where community differences persisted, but sample dispersion was not statistically significant. The NMDS ordination showed a good model fit, and visualizations were generally consistent with the dispersion analyses (Supplementary Figures S4A and S4B). However, when restricting the analysis to samples from *Amphisphaeriales*, *Diaporthales*, and *Glomerellales* (“balanced dataset”), no significant differences were observed by order or family, and the communities were homogeneous across these groups. Although the NMDS showed a similarly good fit, clear clustering was not observed in the visualizations (Supplementary Figures S5A and S5B).

In contrast, the read-based analysis of the full dataset did not detect significant differences in community composition across either fungal orders or families, and

**Table 1** Summary of community structure, composition, and phylogenetic signal analyses across full and balanced datasets using contig- and read-based approaches\*

Analysis	Full dataset**		Balanced dataset**	
Bacteria metagenomics	Contigs	Reads	Contigs	Reads
PERMANOVA (fungal order)	$F_{7,11} = 1.13$ ; $P = \mathbf{0.00008013^*}$	$F_{7,11} = 0.775$ ; $P = 0.5399$	$F_{2,9} = 1.06$ ; $P = 0.312$	$F_{2,9} = 2.3$ ; $P = \mathbf{0.0297^*}$
PERMANOVA (fungal family)	$F_{9,9} = 1.8$ ; $P = \mathbf{0.0044^*}$	$F_{9,9} = 0.6357$ ; $P = 0.7321$	$F_{2,9} = 1.15$ ; $P = 0.1166$	$F_{2,9} = 2.3$ ; $P = \mathbf{0.0311^*}$
PERMUTEST (fungal order)	$F_{7,11} = 1.13$ ; $P = \mathbf{0.0238^*}$	$F_{7,11} = 0.8389$ ; $P = 0.5591$	$F_{2,9} = 0.2678$ ; $P = 0.8776$	$F_{2,9} = 0.8014$ ; $P = 0.4229$
PERMUTEST (fungal family)	$F_{9,9} = 1.8$ ; $P = 0.0907$	$F_{9,9} = 0.6084$ ; $P = 0.7151$	$F_{2,9} = 0.389$ ; $P = 0.7532$	$F_{2,9} = 0.8014$ ; $P = 0.4931$
NMDS stress value	0.0838	0.0000478	0.0829	0.0000806
Phylogenetic signal	$P = \mathbf{0.0026^*}$ ; Mantel $r = 0.955$	$P = \mathbf{0.0000999^*}$ ; Mantel $r = 0.959$	$P = 0.2012$ ; Mantel $r = 0.956$	$P = \mathbf{0.0000999^*}$ ; Mantel $r = 0.959$
Procrustes	$P = \mathbf{0.042^*}$ ; correlation = 0.4689	$P = \mathbf{0.042^*}$ ; correlation = 0.619	$P = \mathbf{0.001^*}$ ; correlation = 0.6871	$P = \mathbf{0.001^*}$ ; correlation = 0.988
Multilevel pattern analysis (indicator species)			<i>Amphisphaeriales</i> vs. <i>Glomerellales</i> = 2 taxa for <i>Glomerellales</i> ( $P = 0.0299$ )	<i>Amphisphaeriales</i> vs. <i>Glomerellales</i> = 37 vs. 102 taxa ( $P = 0.0269$ )
			<i>Diaporthales</i> vs. <i>Glomerellales</i> = 2 taxa for <i>Glomerellales</i> ( $P = 0.0299$ )	<i>Diaporthales</i> vs. <i>Glomerellales</i> = 4 vs. 11 taxa ( $P = 0.0271$ )
			<i>Amphisphaeriales</i> vs. <i>Diaporthales</i> = 1 taxon for <i>Amphisphaeriales</i> ( $P = 0.0287$ )	<i>Amphisphaeriales</i> vs. <i>Diaporthales</i> = 37 vs. 102 taxa ( $P = 0.0269$ )
Multinomial classification			<i>Diaporthales</i> vs. <i>Glomerellales</i> : 26 generalists, 458 too rare, 1 specialist for <i>Diaporthales</i> . 1 specialist for <i>Glomerellales</i> . <i>Amphisphaeriales</i> vs. <i>Diaporthales</i> : 29 generalists, 475 too rare. <i>Amphisphaeriales</i> vs. <i>Glomerellales</i> : 28 generalists, 2 specialists, 521 too rare for <i>Amphisphaeriales</i> . 2 specialists for <i>Glomerellales</i> .	> 300 "too rare" for all the comparisons
Viral metagenomics	Contigs	Reads	Contigs	Reads
PERMANOVA (fungal order)	$F_{7,11} = 1.5$ ; $P = \mathbf{0.0829^*}$	$F_{7,11} = 0.8659$ ; $P = 0.4317$	$F_{2,9} = 1.55$ ; $P = 0.0967$	$F_{2,9} = 3.52$ ; $P = \mathbf{0.0045^*}$
PERMANOVA (fungal family)	$F_{9,9} = 2.4$ ; $P = 0.0935$	$F_{9,9} = 0.7914$ ; $P = 0.5209$	$F_{2,9} = 1.55$ ; $P = 0.1026$	$F_{2,9} = 3.52$ ; $P = \mathbf{0.0051^*}$
PERMUTEST (fungal order)	$F_{7,11} = 1.5$ ; $P = 0.1038$	$F_{7,11} = 1.31$ ; $P = 0.3283$	$F_{2,9} = 1.92$ ; $P = 0.2104$	$F_{2,9} = 1.4$ ; $P = 0.2827$
PERMUTEST (fungal family)	$F_{9,9} = 2.4$ ; $P = 0.1552$	$F_{9,9} = 0.9092$ ; $P = 0.5019$	$F_{2,9} = 1.92$ ; $P = 0.199$	$F_{2,9} = 1.4$ ; $P = 0.2893$
NMDS stress value	0.0956	0.0000834	0.0607	0.00008005
Phylogenetic signal	$P = 0.3729$ ; Mantel $r = 0.934$	$P = \mathbf{0.0097^*}$ ; Mantel $r = 0.956$	$P = \mathbf{0.0016^*}$ ; Mantel $r = 0.95$	$P = \mathbf{0.00019^*}$ ; Mantel $r = 0.947$
Procrustes	$P = \mathbf{0.001^*}$ ; correlation = 0.8033	$P = \mathbf{0.025^*}$ ; correlation = 0.6974	$P = \mathbf{0.001^*}$ ; correlation = 0.9487	$P = \mathbf{0.001^*}$ ; correlation = 0.991
Multilevel pattern analysis (indicator species)			<i>Amphisphaeriales</i> vs. <i>Glomerellales</i> = no significant taxa	<i>Amphisphaeriales</i> vs. <i>Glomerellales</i> = 39 vs. 36 taxa ( $P = 0.028$ )
			<i>Diaporthales</i> vs. <i>Glomerellales</i> = no significant taxa	<i>Diaporthales</i> vs. <i>Glomerellales</i> = 1 vs. 1 taxa ( $P = 0.0301$ )
			<i>Amphisphaeriales</i> vs. <i>Diaporthales</i> = 2 vs. 0 taxa ( $P = 0.0268$ )	<i>Amphisphaeriales</i> vs. <i>Diaporthales</i> = 1 vs. 7 taxa ( $P = 0.0296$ )

**Table 1** (continued)

Analysis	Full dataset**		Balanced dataset**	
Bacteria metagenomics	Contigs	Reads	Contigs	Reads
Multinomial classification			<i>Diaporthales</i> vs. <i>Glomerellales</i> : 13 generalists, 39 too rare <i>Amphisphaeriales</i> vs. <i>Glomerellales</i> : 13 generalists, 42 too rare <i>Amphisphaeriales</i> vs. <i>Diaporthales</i> : 14 generalists, 38 too rare	> 300 “too rare” for all the comparisons
Viral metatranscriptomics	Contigs	Reads	Contigs	Reads
PERMANOVA (fungal order)	$F_{7,11} = 9.48$ ; $P = \mathbf{0.0006^*}$	$F_{7,11} = 0.66$ ; $P = 0.9516$	$F_{2,9} = 1.87$ ; $P = \mathbf{0.0259^*}$	$F_{2,9} = 1.39$ ; $P = 0.0979$
PERMANOVA (fungal family)	$F_{9,9} = 10.8$ ; $P = \mathbf{0.0007^*}$	$F_{9,9} = 1.66$ ; $P = 0.4005$	$F_{2,9} = 1.06$ ; $P = 0.3807$	$F_{2,9} = 1.39$ ; $P = 0.0967$
PERMUTEST (fungal order)	$F_{7,11} = 13.9$ ; $P = \mathbf{0.0002^*}$	$F_{7,11} = 2.18$ ; $P = 0.1818$	$F_{2,9} = 3.74$ ; $P = \mathbf{0.0557^*}$	$F_{2,9} = 0.38$ ; $P = 0.6943$
PERMUTEST (fungal family)	$F_{9,9} = 10.87$ ; $P = \mathbf{0.0012^*}$	$F_{9,9} = 2.3$ ; $P = 0.1342$	$F_{2,9} = 9.03$ ; $P = \mathbf{0.0041^*}$	$F_{2,9} = 1.39$ ; $P = 0.0967$
NMDS stress value	0.0902	0.081	0.0621	0.0914
Phylogenetic signal	$P = 0.3999$ ; Mantel $r = 0.959$	$P = 0.6821$ ; Mantel $r = 0.932$	$P = 0.0539^*$ ; Mantel $r = 0.959$	$P = 0.445$ ; Mantel $r = 0.95$
Procrustes	$P = \mathbf{0.001^*}$ ; correlation = 0.8033	$P = \mathbf{0.001^*}$ ; cor- relation = 0.7716	$P = \mathbf{0.001^*}$ ; correlation = 0.951	$P = \mathbf{0.001^*}$ ; correlation = 0.9689
Multilevel pattern analysis (indicator species)			<i>Amphisphaeriales</i> vs. <i>Glomerellales</i> = 0 vs. 2 ( $P = 0.029$ )	<i>Amphisphaeriales</i> vs. <i>Glomerellales</i> = 24 vs. 19 taxa ( $P = 0.03$ )
			<i>Diaporthales</i> vs. <i>Glomerellales</i> = 0 vs. 1 ( $P = 0.0268$ )	<i>Diaporthales</i> vs. <i>Glomerellales</i> = 18 vs. 2 taxa ( $P = 0.0307$ )
			<i>Amphisphaeriales</i> vs. <i>Diaporthales</i> = no significant taxa	<i>Amphisphaeriales</i> vs. <i>Diaporthales</i> = 10 vs. 5 taxa ( $P = 0.0285$ )
Multinomial classification			<i>Diaporthales</i> vs. <i>Glomerellales</i> : 64 generalists, 272 too rare, 4 specialists for <i>Diaporthales</i> . 9 specialists for <i>Glomerellales</i> : <i>Amphisphaeriales</i> vs. <i>Glomerellales</i> : 80 generalists, 266 too rare, 3 specialists for <i>Glomerellales</i> <i>Amphisphaeriales</i> vs. <i>Diaporthales</i> : 85 generalists, 255 too rare, 4 specialists for <i>Diaporthales</i>	> 300 “too rare” for all the comparisons

\*Complete results are in Supplementary Figures S4–S7, S9–S12, S15–S20, and Supplementary Tables S11–S13, S19, S20, S26–S28

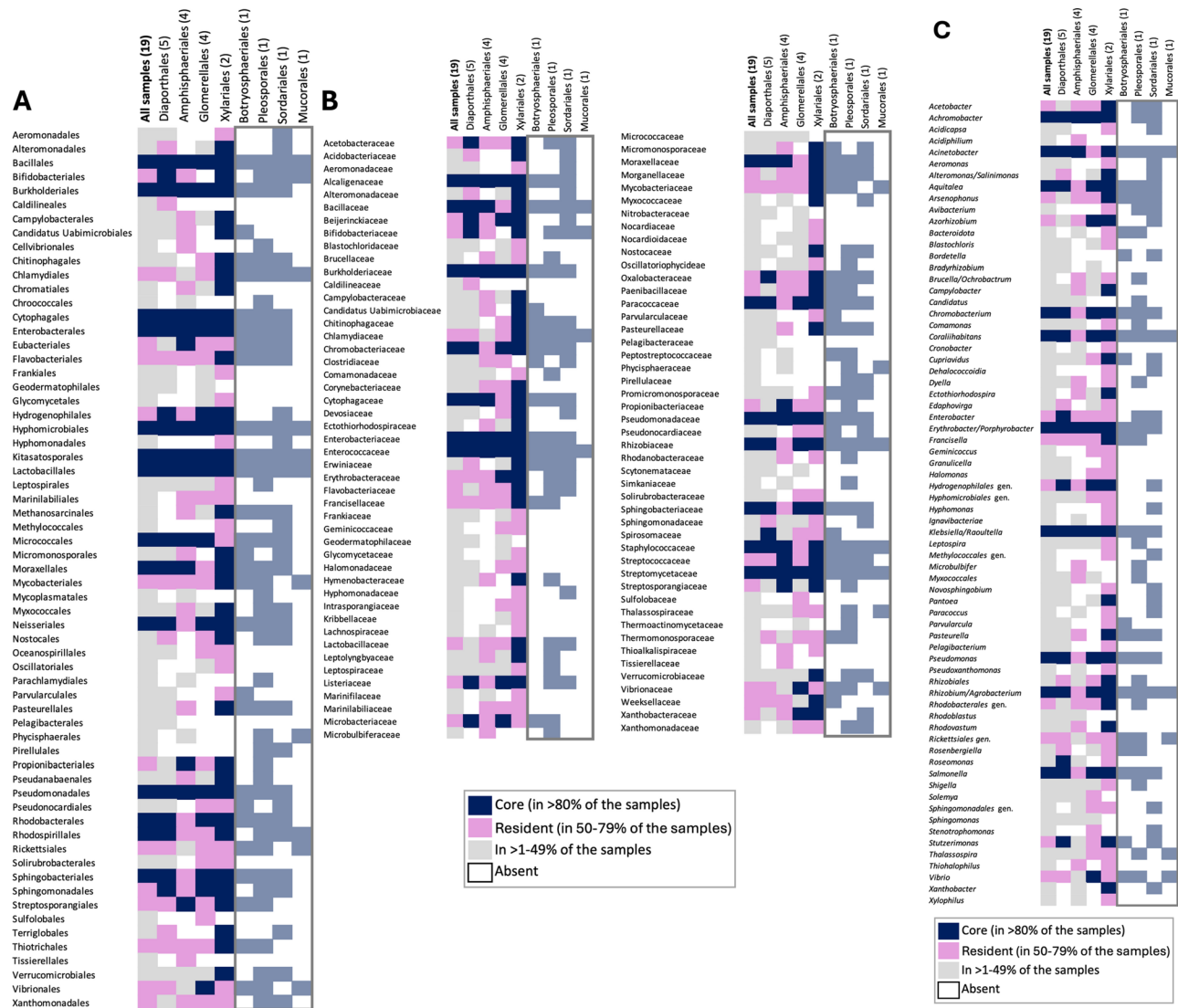
\*\*Full dataset: All samples. Balanced dataset: Fungal orders or families with > 3 samples, i.e., *Amphisphaeriales* (*Pestalotiopsis* spp.), *Diaporthales* (*Diaporthaceae*), and *Glomerellales* (*Glomerellaceae*)

group dispersions were also not significant. The NMDS analysis produced a very low stress value, likely reflecting the minimal variation in the dataset, and no distinct clustering patterns were visible in the ordinations (Supplementary Figures S6A and S6B). However, dbRDA visualizations showed some tendencies toward clustering for *Diaporthales* and *Glomerellales* at the order level (Supplementary Figure S6C), and for *Glomerellaceae* at the family level (Supplementary Figure S6D). Notably, when restricting the analysis to the balanced dataset, the read-based approach did detect significant differences among communities by both order and family, with homogeneity in dispersion across groups. Again, NMDS

ordination showed a low stress value but no clear clusters, while dbRDA continued to suggest subtle grouping patterns (Supplementary Figure S7).

When using contigs, the bacterial community composition revealed that 14 orders, 14 families, and 10 genera were consistently present in 80–100% of the fungal samples (i.e., core); and 13 orders, 20 families, and 10 genera were present in 50–79% of the samples (i.e., resident) (Fig. 2, Supplementary Tables S4–S7). Remarkably, ~48%, 51%, and 70% of the detected bacteria at the order, family, and genus levels, respectively, were unclassified. The most abundant core bacterial orders across all samples were *Bacillales*, *Burkholderiales*, *Cytophagales*,



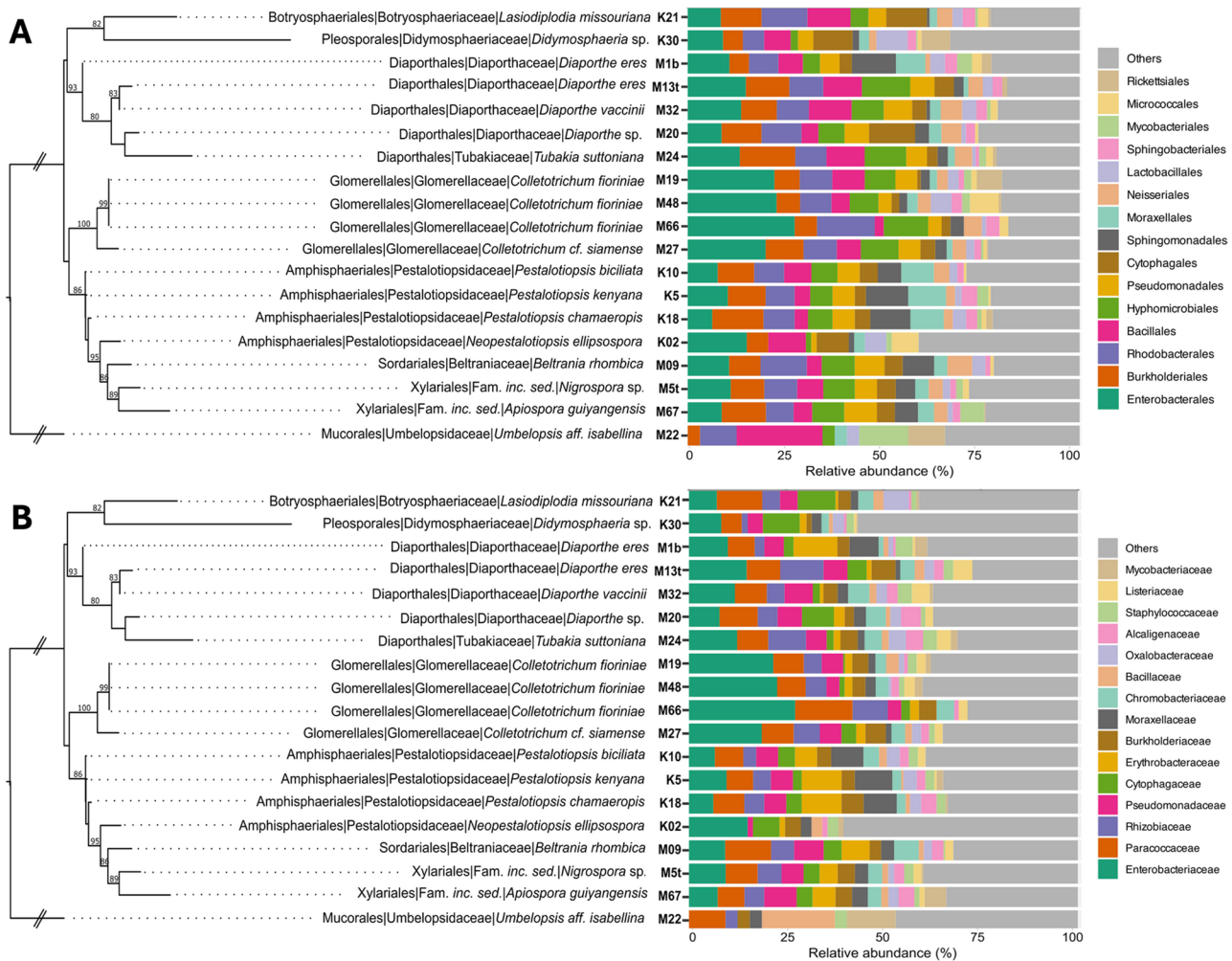


**Fig. 2** Core and resident bacterial orders (A), families (B), and genera (C) by fungal order, for contig data

*Enterobacterales*, *Hyphomicrobiales*, *Lactobacillales*, *Moraxellales*, *Neisseriales*, *Pseudomonadales*, and *Rhodobacterales*, with relative abundances ~1.5–7% (Figs. 2 and 3, Supplementary Tables S4–S7). The top five most abundant core bacterial families and genera were *Alcaligenaceae*, *Moraxellaceae*, *Rhizobiaceae*, *Sphingobacteriaceae*, and *Streptomycetaceae*; and *Coraliihabitans*, *Erythrobacter/Parphyrobacter* complex, *Pseudomonas*, *Rhizobium/Agrobacterium* complex, and *Salmonella*, respectively (Figs. 2 and 3, Supplementary Tables S4–S7, Supplementary Figure S8). The top five most abundant resident bacterial orders, families, and genera across all fungal samples were *Flavobacteriales*, *Mycobacteriales*, *Propionibacteriales*, *Rickettsiales*, and *Sphingomonadales*; *Beijerinckiaceae*, *Lactobacillaceae*, *Microbacteriaceae*, *Morganellaceae*, and *Streptococcaceae*; and *Azorhizobium*, *Enterobacter*, *Francisella*, unidentified

*Rickettsiales*, and *Vibrio*, respectively (Figs. 2 and 3, Supplementary Tables S4–S7). When using read data, no core orders, families, or genera were detected across all samples, and only, *Kitasatosporales* and *Streptomycetaceae* were classified as resident (Fig. 4, Supplementary Tables S8–S10). Interestingly, only a few reads, 0.02% and 0.75%, of the detected prokaryotes at the order and family levels were unclassified, in contrast to 92% unclassified at the genus level. For example, *Nostoc* and *Clostridioides* were some of the most abundant genera that were classified at that level (Fig. 5, Supplementary Table S3). Most other abundant genera were classified only at the family or order level (Fig. 5, Supplementary Table S3).

Using contig data, only *Amphisphaerales*, *Diaporthales*, and *Glomerellales*, had four or more samples each to make inferences about core or resident taxa at the fungal order level (Fig. 2, Supplementary Tables S5–S7).

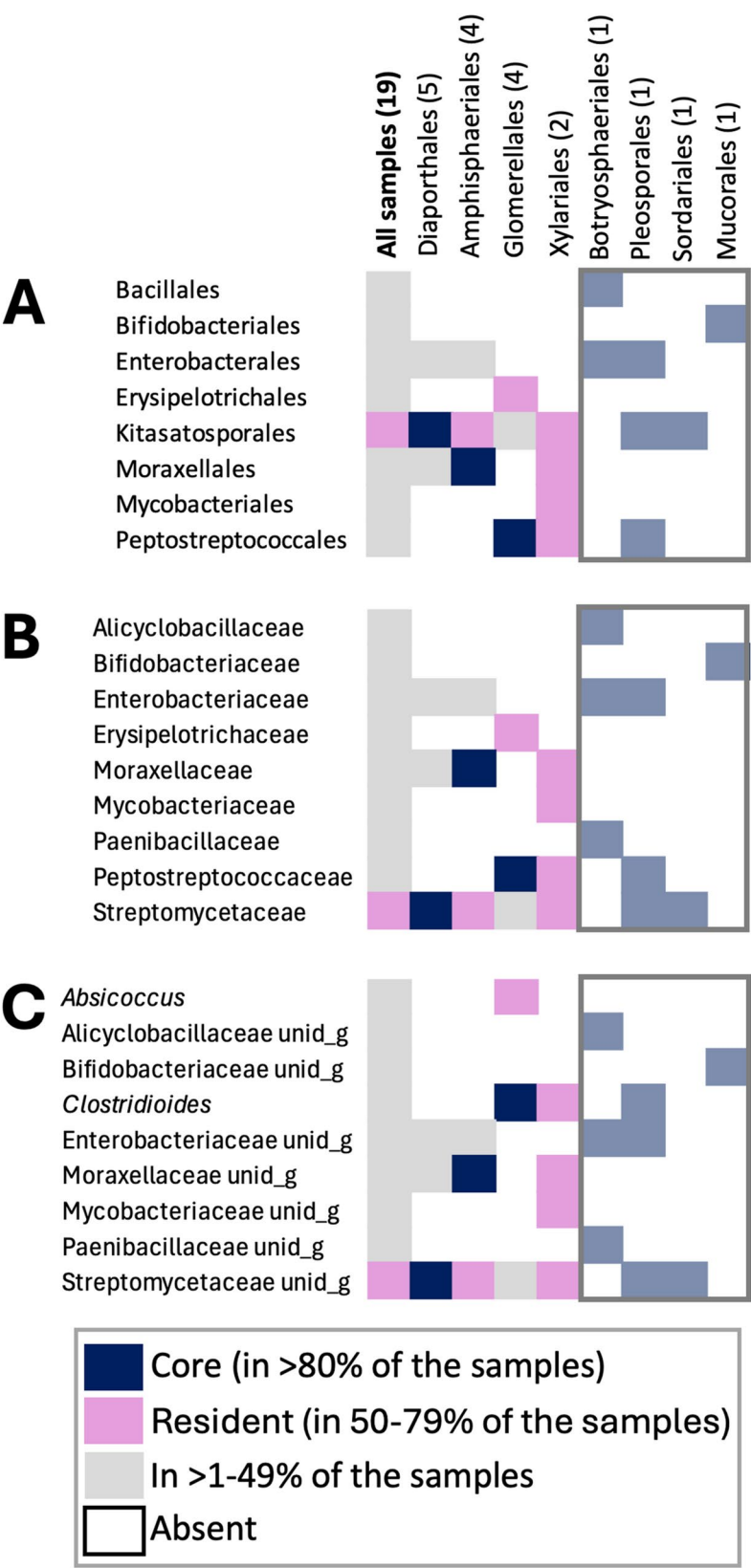


**Fig. 3** Relative abundance (%) of the top 15 most abundant bacterial orders (A) and families (B) among the fungal isolates for contig data. The Maximum Likelihood phylogenetic tree was constructed using concatenated nrDNA ITS, TEF, and TUB sequences. Bootstrap values are shown on nodes. Unidentified or unclassified taxa were excluded

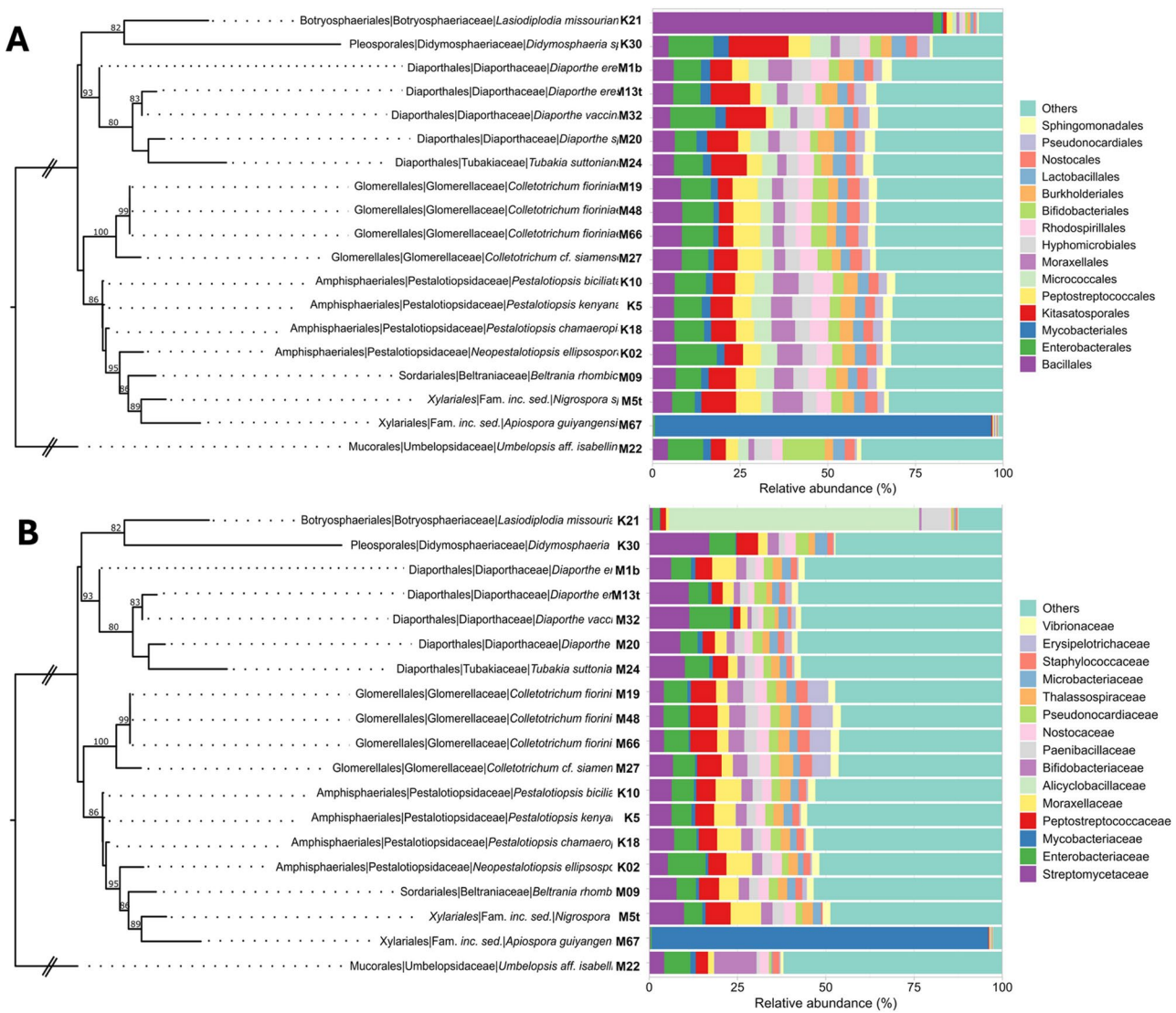
The *Amphisphaeriales* had 13 orders, 13 families, and 4 genera classified as core taxa. Examples of core taxa that also had high relative abundances include *Bacillales*, *Burkholderiales*, *Hyphomicrobiales*, and *Lactobacillales*; *Acetobacteraceae*, *Bacillaceae*, *Burkholderiaceae*, and *Enterobacteriaceae*; and *Achromobacter*, *Acinetobacter*, *Erythrobacter*/*Porphyrobacter*, and *Klebsiella*/*Raoultella* (Figs. 2 and 3, Supplementary Tables S4–S7). Using read data, only *Moraxellaceae* (*Moraxellales*) were classified as core (Fig. 4, Supplementary Tables S8–S10). The fungal order *Diaporthales* had 17 orders, 21 families, and 14 genera classified as core. Examples of those taxa are *Bacillales*, *Burkholderiales*, *Cytophagales*, and *Hyphomicrobiales*; *Acetobacteraceae*, *Bacillaceae*, *Enterobacteraceae*, and *Rhizobiaceae*; and *Coraliihabitans*, *Erythrobacter*/*Porphyrobacter*, *Pseudomonas*, and *Rhizobium*/*Agrobacterium*. For *Diaporthales*, only *Kitasatosporales* and *Streptomycetaceae* were classified as core using read data. Lastly, *Glomerellales* samples had 15 orders, 17 families,

and 12 genera classified as core. Examples of core taxa with high relative abundances include *Burkholderiales*, *Enterobacterales*, *Hyphomicrobiales*, and *Rhodobacterales*; *Enterobacteriaceae*, *Pseudomonadaceae*, *Paracoccaceae*, and *Rhizobiaceae*; and *Coraliihabitans*, *Klebsiella*/*Raoultella*, *Pseudomonas*, and *Rhizobium*/*Agrobacterium* (Figs. 4 and 5, Supplementary Tables S3, S8–S10). Only *Clostridioides* (*Peptostreptococcales*, *Peptostreptococcaceae*) was classified as core using read data.

Analyses of indicator and specialist taxa revealed differences between the contig- and read-based approaches. A summary of the results, including p-values and other statistics, is provided in Table 1. Using contigs, multilevel pattern analysis identified a limited number of significant associations. For instance, the bacterial family *Streptosporangiaceae* was associated with *Amphisphaeriales* but not with *Diaporthales* or *Glomerellales*. For *Glomerellales*, two bacterial families (*Rhizobiaceae* and *Enterobacteriaceae*) showed significant associations compared



**Fig. 4** Core and resident bacterial orders (A), families (B), and genera (C) by fungal order, for read data



**Fig. 5** Relative abundance (%) of the top 15 most abundant bacterial orders (A) and families (B) among the fungal isolates for read data. The Maximum Likelihood phylogenetic tree was constructed using concatenated nrDNA ITS, TEF, and TUB sequences. Bootstrap values are shown on nodes. Unidentified or unclassified taxa were excluded

to both *Amphisphaerales* and *Diaporthales*. The multinomial species test further identified several potential specialists: *Moraxellales* and *Sphingomonadales* were associated with *Amphisphaerales*, while *Enterobacteriales* and *Rickettsiales* were associated with *Glomerellales* (Supplementary Figure S9). *Diaporthales* had only one specialist (*Cytophagales*), whereas *Glomerellales* had three (*Enterobacteriales*, *Hyphomicrobiales*, and *Micrococcales*) (Supplementary Figure S10). The comparison between *Amphisphaerales* and *Diaporthales* yielded no specialist taxa but revealed generalist and rare bacterial groups, including *Burkholderiales*, *Enterobacteriales*, and *Pseudomonadales* (Supplementary Figure S11).

In contrast, the read-based approach, especially using the balanced dataset, identified more numerous and

taxonomically diverse significant associations. The results from the multinomial classification test for the best-represented fungal orders demonstrated a hypersensitivity due to the abundant taxa with zeros and low values, as more than 300 taxa were classified as “too rare” [74]. In the multilevel pattern analysis, the comparison between *Amphisphaerales* and *Diaporthales* yielded 37 significant indicator bacterial taxa for *Amphisphaerales*—mainly from *Actinomycetota*, *Bacillota*, and *Pseudomonadota*—and 102 taxa for *Diaporthales*, enriched in *Actinomycetota*, *Bacteroidota*, and *Pseudomonadota* (Supplementary Table S11). Similar results were observed in the comparison between *Amphisphaerales* and *Glomerellales*, with 37 significant taxa in the former and 103 in the latter, largely from *Actinomycetota* and



*Pseudomonadota* (Supplementary Table S12). The *Diaporthales* vs. *Glomerellales* comparison resulted in only a few indicator taxa: 4 for *Diaporthales* and 11 for *Glomerellales* (Supplementary Table S13). Besides the number of indicator taxa, the main difference with the contig-based approach is that most of the taxa in the read-based approach were not identified to the genus level.

Both contig and read-based approaches detected a significant phylogenetic signal in the endohyphal bacterial communities (Table 1). The contig-based Mantel test yielded a *P*-value of 0.0026 and a robust model fit based on 10,000 permutations. Procrustes analysis showed marginal significance with a moderate correlation (Table 1, Supplementary Figure S12A). The read-based Mantel test produced a strong signal and model fit. The Procrustes analysis showed stronger congruence than the contig-based approach (Supplementary Figure S6E). Notably, with the balanced dataset, the significant signal remained, and the Procrustes test showed even greater alignment (Supplementary Figure S7E), suggesting stronger phylogenetic structuring among bacterial communities when fungal taxa are taxonomically balanced.

## Viral metagenomic data analyses

### Alpha diversity

A total of 851,000 reads matched 1,008 viral taxa. On average, each sample contained approximately 200 taxa, with M67 and K21 exhibiting the highest richness (479 and 324 taxa, respectively). In contrast, M9 and M66 had the lowest richness, each with 171 taxa. Although sample M67 yielded nearly 200,000 reads—similar to the other samples—its rarefaction curve did not reach an asymptote (Supplementary Figure S13A). At the order level, *Amphisphaeriales* harbored the greatest number of viral taxa (658), followed by *Botryosphaeriales*, *Diaporthales*, and *Glomerellales*, each with approximately 300 taxa. *Sordariales* had 171 taxa, while *Mucorales*, *Pleosporales*, and *Xylariales* each recorded around 200. None of the orders appeared to reach an asymptote, and extrapolation suggested that additional taxa may still be recovered (Supplementary Figure S13B). Family-level analysis showed that *Apiosporaceae* had the highest viral richness (479 taxa), followed closely by *Pestalotiopsidaceae* (397). *Beltraniaceae* and *Tubakiaceae* recorded the fewest taxa, with fewer than 200 each, while the remaining families ranged between 200 and 380 taxa (Supplementary Table S14). Among all families, only *Apiosporaceae* showed a rarefaction curve that approached an asymptote (Supplementary Figure S13C).

After assembly and taxonomic classification, metagenomic data yielded 978 contigs that matched viruses (Supplementary Table S15). Sample comparison revealed significant differences in both the number of taxa and contigs. Samples K21 and M67 exhibited similar values,

with around 70 taxa and 100 contigs each, while other samples had fewer than 100 contigs and approximately 50 taxa (Supplementary Figure S14A). When classified by fungal host order, *Pleosporales*, *Sordariales*, and *Xylariales* had similar contig counts (~200), though *Xylariales* displayed the highest taxonomic diversity with around 80 taxa. The taxonomic composition of *Botryosphaeriales* was comparable to these groups, while other orders showed fewer than 50 taxa and 100 contigs. Notably, *Glomerellales*, *Diaporthales*, and *Sordariales* appeared to approach an asymptote (Supplementary Figure S14B). At the family level, samples M5t and M67 accounted for about 200 contigs and nearly 100 taxa. *Botryosphaeriaceae* followed, with around 100 contigs and a slightly lower taxonomic count. Other families contained fewer than 50 taxa. The asymptote was reached by *Diaporthaceae*, *Glomerellaceae*, and *Pestalotiopsidaceae* (Supplementary Figure S14C).

### Viral community structure, composition, and phylogenetic signal across fungal hosts (metagenomics)

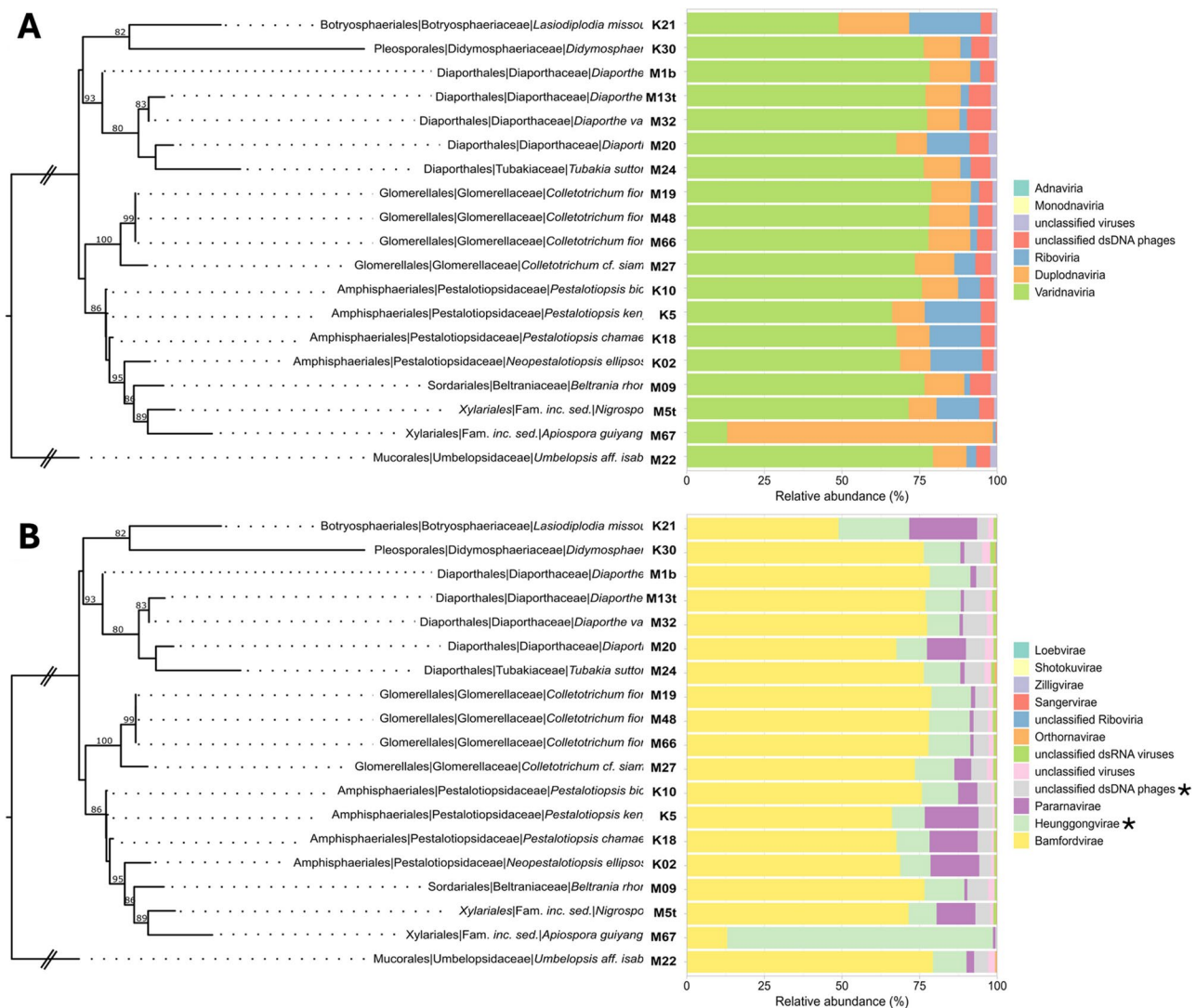
Analyses of viral community structure yielded somewhat distinct results depending on whether contigs or reads were used (Table 1). Using the contig-based approach, the analysis of the complete dataset revealed no significant differences in viral community composition across fungal orders or families. Community dispersion was also statistically homogeneous for orders and families. The NMDS ordination produced a stress value of 0.0956, indicating a good model fit, with most samples clustering in the lower-left quadrant (Supplementary Figures S15A and S15B). When restricting the analysis to the most taxonomically represented fungal orders (balanced dataset), the results remained non-significant for both community composition and dispersion, for orders and families. Nonetheless, the NMDS stress value for this balanced dataset was 0.0607, still within the acceptable range, and visualizations revealed modest clustering, with *Amphisphaeriales* forming the most consistent group, while a few samples from *Diaporthales* and *Glomerellales* showed more scattered patterns (Supplementary Figures S15C and S15D).

The read-based analysis of the complete dataset also found no significant differences in viral communities when classified by fungal order or family, and homogeneity of dispersion was likewise observed (Table 1). The NMDS analysis resulted in a very low stress value, suggesting limited variation or sparse data. No clear clustering was observed in the ordinations (Supplementary Figures S16A and S16B). However, the dbRDA analysis showed some grouping patterns among the best-represented orders and families (Supplementary Figures S16C and S16B). When the analysis was restricted to the balanced dataset, the read-based approach did detect

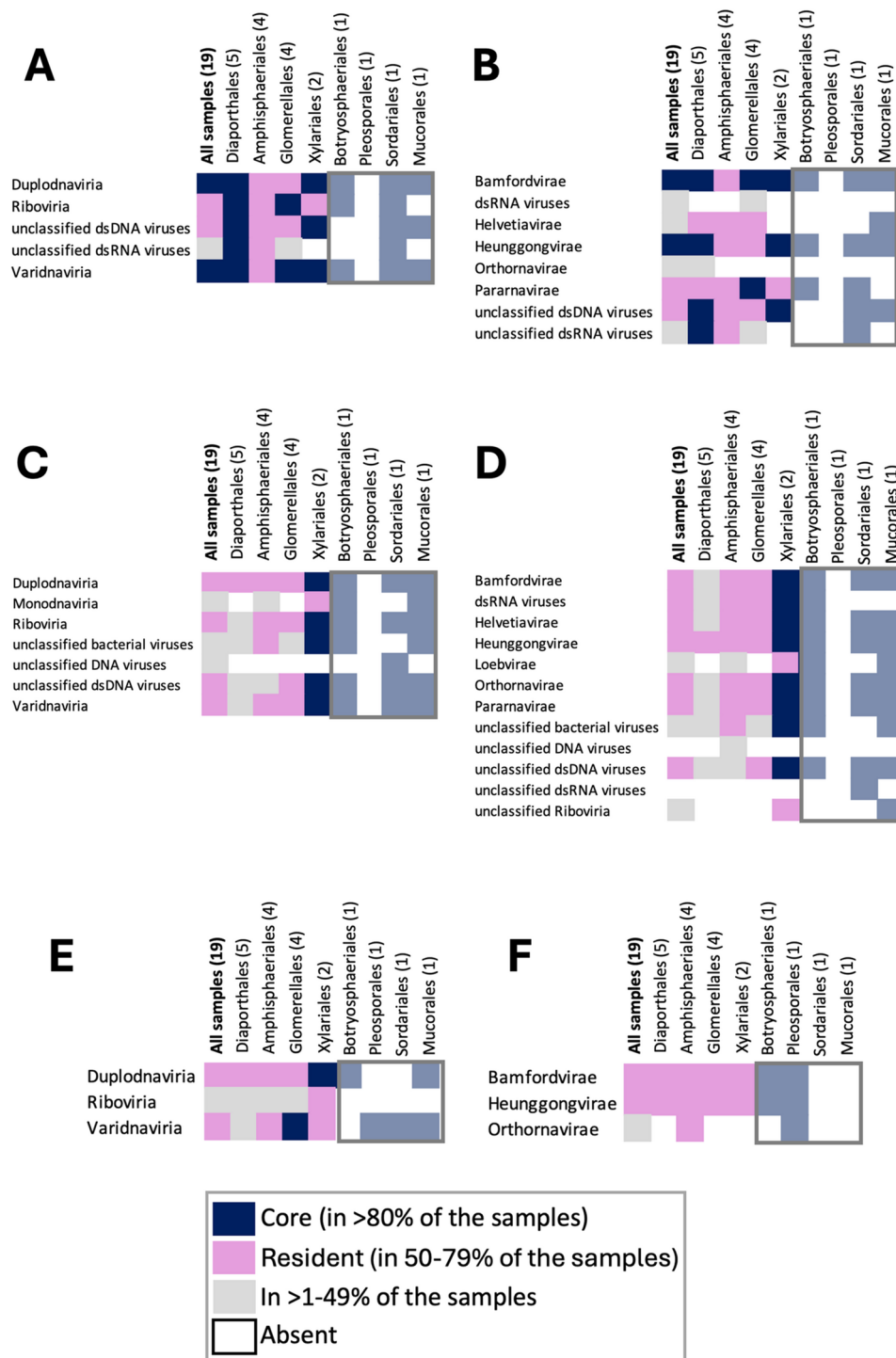
significant differences in viral communities by both fungal order and family, while community dispersion remained homogeneous. Despite a similarly low NMDS stress value, no clear clustering was observed in ordination plots (Supplementary Figures S17A and S17B). Nevertheless, dbRDA visualizations again revealed separation patterns aligned with fungal taxonomy, except for sample M1b (*Diaporthales*), which clustered apart from other *Diaporthales* samples (Supplementary Figures S17C and S17D).

For read-based data, only one taxon (*Duplodnaviria*, *Heunggongvirae*, *Caudoviricetes*) for one sample (M67) had a relative abundance of more than 0.1% (Supplementary Table S16). Therefore, inferences regarding core or resident taxa were not made. Figure 6 illustrates overall relative abundances at the realm and kingdom levels for

read data (see also Supplementary Table S14). The viral community composition using contig data revealed that 2 realms (*Duplodnaviria* and *Varidnaviria*) and 2 kingdoms (*Bamfordvirae* and *Heunggongvirae*) were consistently present in 80–100% of the fungal samples (i.e., core); and 2 realms (*Riboviria* and unclassified dsDNA viruses) and 2 kingdoms (unclassified dsDNA viruses and *Pararnavirae*) were present in 50–79% of the samples (i.e., resident) (Fig. 7, Supplementary Tables S17 and S18). Interestingly, only about 5% of the detected viral sequences at the realm and kingdom levels were unclassified (Fig. 8, Supplementary Table S15). The realms *Duplodnaviria* and *Varidnaviria*, and kingdoms *Bamfordvirae* and *Heunggongvirae* were also the most abundant, with relative abundances ranging from ~20–60% (Fig. 8, Supplementary Table S15).



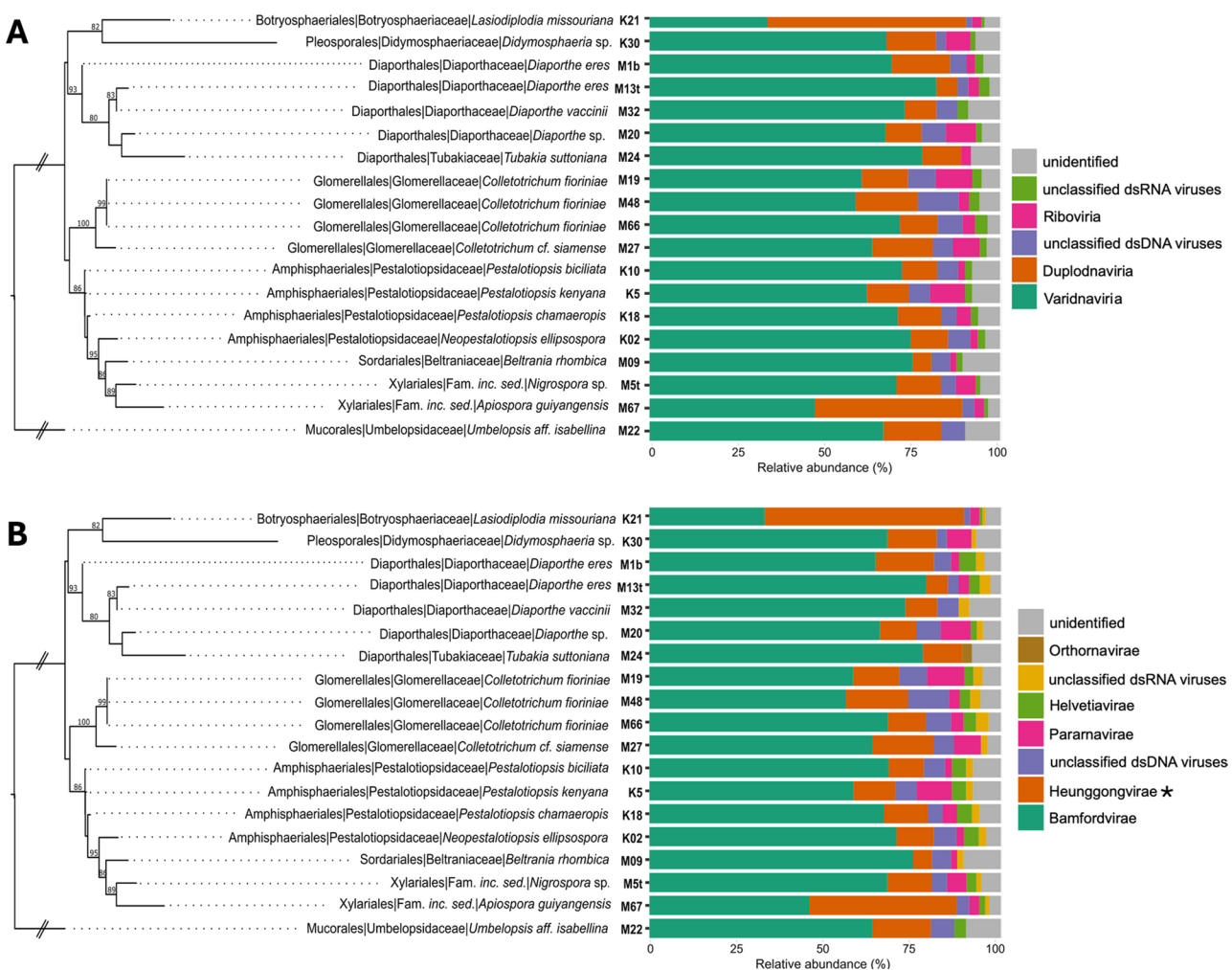
**Fig. 6** Relative abundance (%) of the viral realms (A) and kingdoms (B) among the fungal isolates, using metagenomic read data. The Maximum Likelihood phylogenetic tree was constructed using concatenated nrDNA ITS, TEF, and TUB sequences. Bootstrap values are shown on nodes. For B, potential taxa containing bacteriophages are indicated by \*



**Fig. 7** Core and resident viral realms (A, C, E) and kingdoms (B, D, F) by fungal order, for the metagenomic and metatranscriptomic contig and read datasets. A, B. Metagenomic contig data. C, D. Metatranscriptomic contig data. E, F. Metatranscriptomic read data. Metagenomic read data is not included because only one taxon (*Duplodnaviria*, *Heunggongvirae*) for one sample (M67) had a relative abundance of >0.1%

Only *Amphisphaeriales*, *Diaporthales*, and *Glomerellales* had four or more samples each to make inferences about core or resident taxa for contig data (Fig. 7, Supplementary Tables S17 and S18). The *Amphisphaeriales* had no taxa classified as core, but 5 realms and 6 kingdoms

classified as resident. The resident realms, from the most abundant to the least, were *Varidnaviria*, *Duplodnaviria*, unclassified dsDNA viruses, *Riboviria*, and unclassified dsRNA viruses; and kingdoms *Bamfordvirae*, *Heunggongvirae*, unclassified dsDNA viruses, *Pararnavirae*,



**Fig. 8** Relative abundance (%) of the viral realms (A) and kingdoms (B) among the fungal isolates, using metagenomic contig data. The Maximum Likelihood phylogenetic tree was constructed using concatenated nrDNA ITS, TEF, and TUB sequences. Bootstrap values are shown on nodes. For B, potential taxa containing bacteriophages are indicated by \*

*Helvetiavirae*, and unclassified dsRNA viruses (Figs. 7 and 8, Supplementary Tables S15, S17, S18). The fungal order *Diaporthales* had 5 realms and 4 kingdoms classified as core, and no resident realms but 2 resident. The core realms, from the most abundant to the least, were *Varidnaviria*, *Duplodnaviria*, unclassified dsDNA viruses, *Riboviria*, and unclassified dsRNA viruses; and kingdoms *Bamfordvirae*, *Heunggongvirae*, unclassified dsDNA viruses, and unclassified dsRNA viruses (Figs. 7 and 8, Supplementary Tables S15, S17, S18). The resident kingdoms were *Helvetiavirae* and *Paramyovirae*. Lastly, *Glomerellales* had 2 core realms (*Riboviria* and *Varidnaviria*), 2 core kingdoms (*Bamfordvirae* and *Paramyovirae*), 2 resident realms (*Duplodnaviria* and unclassified dsDNA viruses) and 3 resident kingdoms (*Helvetiavirae*, *Heunggongvirae*, and unclassified dsDNA viruses) (Fig. 7, Supplementary Tables S17 and S18).

Analyses of viral indicator taxa and phylogenetic signal showed differences between the contig-based and

read-based approaches (Table 1). Using contigs, the multilevel pattern analysis identified only two viral taxa—*Duplodnaviria* (*Uroviricota*, *Caudoviricetes*) and *Varidnaviria* (*Nucleocytoviricota*, *Megaviricetes*)—as significantly associated with the fungal order *Amphisphaeriales* compared to *Diaporthales*. However, the multinomial species classification test did not detect any specialist taxa among *Amphisphaeriales*, *Diaporthales*, or *Glomerellales*. Instead, all three groups harbored generalist and rare taxa, with *Varidnaviria* and *Duplodnaviria* repeatedly identified as generalists, while unclassified dsRNA viruses and *Riboviria* were frequently categorized as rare (Supplementary Figures S18–S20).

In contrast, the read-based approach detected a broader range of viral associations, though interpretation was limited by the high number of low-abundance or “too rare” taxa. In the comparison between *Amphisphaeriales* and *Diaporthales*, only one significant taxon (*Riboviria*, *Paramyovirae*, *Retraviricetes*) was associated



with *Amphisphaeriales*. *Diaporthales*, on the other hand, was characterized by *Megaviricetes*, *Naldaviricetes*, and *Pokkesviricetes* (*Varidnaviria*, *Bamfordvirae*) (Supplementary Table S19). The comparison between *Amphisphaeriales* and *Glomerellales* yielded more informative results, identifying over 30 significant viral taxa for each group, primarily belonging to *Varidnaviria* and *Duplodnaviria* (Supplementary Table S20). The *Diaporthales*–*Glomerellales* comparison resulted in only one notable taxon per group: *Caudoviricetes* (*Duplodnaviria*) for *Diaporthales* and *Megaviricetes* (*Varidnaviria*) for *Glomerellales*.

Both approaches detected evidence of phylogenetic signal, though the strength and consistency differed (Table 1). Using contigs, the analysis of the full dataset showed no significant phylogenetic signal. However, when restricted to the best-represented fungal orders (*Amphisphaeriales*, *Diaporthales*, *Glomerellales*), the Mantel test indicated a significant signal and strong model fit. The Procrustes test further confirmed correlation between viral community structure and host phylogeny (Table 1, Supplementary Figure S12B). The read-based analysis revealed a significant phylogenetic signal even in the full dataset (Table 1). This result was further strengthened in the balanced dataset. The Procrustes analysis also showed significant alignment and correlation between viral and fungal phylogenies (Table 1, Supplementary Figure S16E), and a much stronger result in the balanced dataset (Supplementary Figure S17E).

### Viral metatranscriptomic data analyses

#### Alpha diversity

A total of 7,385,922 reads matched 1,373 viral taxa following taxonomic classification. The samples with the highest richness were M1b and M5t, each with just under 700 taxa. Samples K21, K5, M19, M20, and M67 each contained over 600 taxa, while the remaining samples ranged between 300 and 500 taxa (Supplementary Table S21). Most samples appeared to approach an asymptote in rarefaction curves, except for M13t (Supplementary Figure S21A). At the order level, *Amphisphaeriales* and *Diaporthales* exhibited the highest viral richness, each with over 900 taxa, followed by *Glomerellales* with 819; these three orders showed a clear tendency to reach an asymptote. *Mucorales* also approached an asymptote with approximately 500 taxa (Supplementary Table S21). In contrast, *Xylariales* registered around 740 taxa, *Botryosphaeriales* approximately 600, *Pleosporales* just over 500, and *Sordariales* the fewest, with around 300 taxa. These orders did not reach an asymptote, and extrapolated estimates remained under one million reads (Supplementary Figure S21B). When analyzed by fungal family, *Diaporthaceae*, *Glomerellaceae*, and *Pestalotiopsidaceae*

each had over 800 viral taxa and showed a strong trend toward reaching an asymptote. *Umbelopsidaceae* also approached an asymptote, with slightly over 500 taxa. *Xylariales incertae sedis* accounted for around 740 taxa, while *Beltraniaceae* had the lowest richness, with 319 taxa. The remaining families ranged between 400 and 540 taxa, none of which reached an asymptote (Supplementary Table S21).

After assembly and taxonomic classification, the meta-transcriptomic data recorded a total of 8,245 contigs that matched viruses (Supplementary Table S22). Most samples contributed over 500 contigs, except for M9, M32, and M48, which had around 200. However, extrapolated data predicted up to 1,000 contigs and nearly 400 taxa. On average, the number of taxa across samples was around 200 (Supplementary Figure S22A). At the fungal order level, *Sordariales* exhibited the highest diversity, with 300 taxa across 1,000–2,500 contigs. *Amphisphaeriales* accounted for fewer than 100 taxa and 1,000 contigs, while other orders presented around 200 taxa and close to 1,000 contigs. Extrapolated data suggested that most groups could reach approximately 400 taxa (Supplementary Figure S22B). At the family level, *Diaporthaceae*, *Glomerellaceae*, and *Pestalotiopsidaceae* showed the highest number of contigs (~1,500) and taxa (>200). Most other families reached around 500 contigs and 200 taxa, while *Beltraniaceae* had the fewest (<1,000 contigs and <100 taxa). Extrapolated data indicated the potential for all groups to have higher reads and taxon counts, with *Diaporthaceae* the only family approaching an asymptote (Supplementary Figure S22C).

#### Viral community structure, composition, and phylogenetic signal across fungal hosts (metatranscriptomics)

Beta diversity analyses produced different outcomes depending on whether viral communities were assessed using contigs or reads (Table 1). Using the contig-based approach, beta diversity analysis of the full dataset revealed significant differences in viral community composition across both fungal orders and families. These differences were accompanied by significant heterogeneity in community dispersion at the fungal order and family levels. The NMDS ordination indicated a good model fit, though no strong clustering patterns were observed for either fungal taxonomic level (Supplementary Figures S23A and S23B). When the analysis was restricted to the balanced dataset, significant differences were still observed at the order level, but not at the family level. Community dispersion was homogeneous across fungal orders, yet heterogeneous across families. NMDS again provided a good fit, and some taxonomic structure was evident—samples from *Glomerellales* (e.g., M19 and M48) clustered together, *Amphisphaeriales* formed a distinct group, and *Diaporthales* samples appeared in two

separate quadrants (Supplementary Figures S23C and S23D).

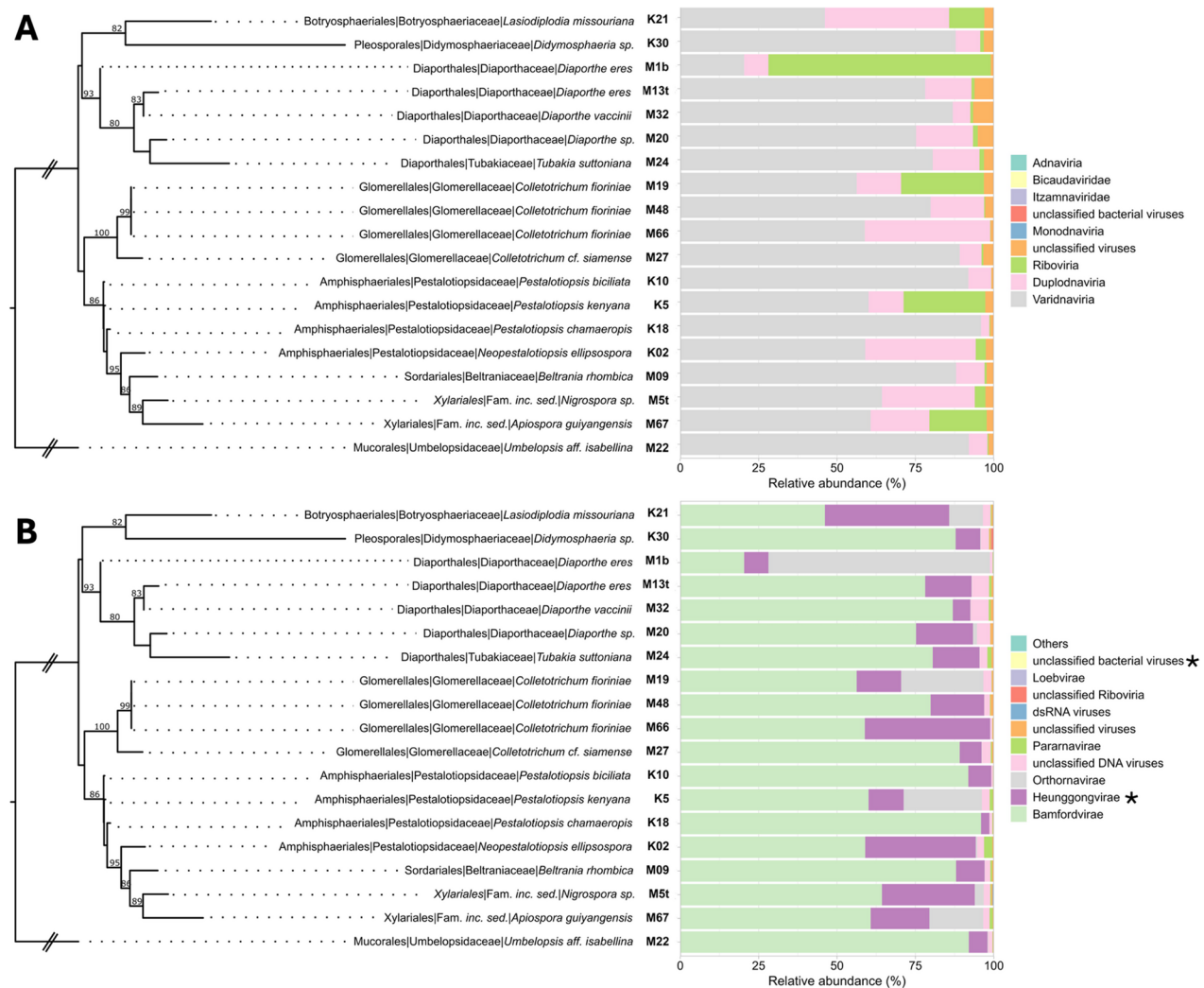
In contrast, the read-based approach did not detect significant differences in viral communities at either the order or family level, for both the full and balanced datasets (Table 1). Dispersion was also statistically homogeneous across comparisons. Nevertheless, the NMDS ordination of the full read-based dataset had a good fit and resolved a three-dimensional solution. This enabled visualization of clustering by fungal order, particularly *Diaporthales* and *Glomerellales*, when using the second and third dimensions (Supplementary Figure S24). Similar patterns were seen at the family level, where *Diaporthaceae*, *Glomerellaceae*, and *Pestalotiopsidaceae* showed grouping in ordination space (Supplementary Figure S25). However, the dbRDA plots were less effective in resolving these patterns, though loose clustering of some *Diaporthales* and *Glomerellales* samples and *Glomerellaceae* members was still apparent (Supplementary Figure S26). When restricted to the balanced dataset, the read-based NMDS yielded a two-dimensional solution with a good fit, but no clear taxonomic clustering emerged. Samples were scattered across axes regardless of their order or family assignments (Supplementary Figure S27). The dbRDA confirmed this result, showing similarly diffuse patterns with no consolidated groupings (Supplementary Figure S27).

The viral community composition using metatranscriptomic contig or read data revealed no realms or kingdoms consistently present in 80–100% of the fungal samples (i.e., core). Using contig data, 4 realms (*Duplodnaviria*, *Riboviria*, *Varidnaviria*, and unclassified dsDNA viruses) and several kingdoms (*Bamfordvirae*, *Helvetiavirae*, *Heunggongvirae*, *Orthornavirae*, *Pararnavirae*, and unclassified dsRNA and dsDNA viruses) were present in 50–79% of the samples (i.e., resident) (Fig. 7, Supplementary Tables S23 and S24). Using read data, two realms (*Duplodnaviria* and *Varidnaviria*) and two kingdoms (*Bamfordvirae* and *Heunggongvirae*) were classified as resident (Fig. 7, Supplementary Table S25). Figure 9 shows relative abundances of viral realms and kingdoms using read data (see also Supplementary Table S21); about 42% of the detected viral sequences at the realm and kingdom levels were unclassified. Using contig data, the realms *Duplodnaviria* and *Varidnaviria*, and kingdoms *Bamfordvirae* and *Heunggongvirae* were also the most abundant, and consistent with the metagenomic data, with relative abundances of approximately 25% (Fig. 10, Supplementary Table S22).

Using read data, *Varidnaviria* was the most abundant realm for all the samples, with the exception of M1b (*Diaporthe eres*) where *Riboviria* was dominant (Supplementary Fig. 9). The read-based approach identified a greater number of *Riboviria* (including dsRNA viruses

and potential mycoviruses) compared to the contig-based approach. For instance, the contig data recovered *Diplodia scrobiculata* RNA virus 1-like (23 contigs across 9 samples) along with 107 additional *Riboviria*-classified contigs. In contrast, the read data yielded 737,538 reads assigned to *Riboviria*, including members of *Alternavirus*, *Scleroulivirus*, and *Tobamovirus*, distributed across 4 samples (Supplementary Tables S21 and S22).

Only *Amphisphaeriales*, *Diaporthales*, and *Glomerellales* had at least four samples each, allowing inferences about core and resident viral taxa. Based on contig data, none of the three fungal orders had viral realms or kingdoms that met the threshold to be classified as core. However, using read data, *Varidnaviria* was identified as a core realm for *Glomerellales* (Fig. 7; Supplementary Tables S23 and S24). For resident taxa, *Amphisphaeriales* harbored three realms and five kingdoms according to contig data, and two realms and three kingdoms based on read data. *Diaporthales* contained one resident realm and one kingdom in the contig dataset, and one realm and two kingdoms in the read dataset (Fig. 7). *Glomerellales* had the highest richness of resident taxa, with three realms and five kingdoms using contig data, and one realm and two kingdoms using read data. In *Amphisphaeriales*, the most abundant resident realm was *Duplodnaviria* (18%) in the contig dataset and *Varidnaviria* (88%) in the read dataset (Figs. 7 and 9, and 10; Supplementary Tables S23–S25). Other realms in the contig data had relative abundances below 0.5%. The most abundant kingdoms in *Amphisphaeriales* (contig data) were *Bamfordvirae* (26%) and *Heunggongvirae* (18%), followed by *Helvetiavirae*, *Pararnavirae*, and various unclassified dsRNA and dsDNA viruses, each below 1.5% (Fig. 9; Supplementary Table S22). In the read dataset, *Bamfordvirae* dominated (88%), with *Heunggongvirae* and *Orthornavirae* each contributing around 6% (Fig. 10, Supplementary Table S21). Approximately 54% of viral contigs from *Amphisphaeriales* could not be classified at the realm or kingdom level, in contrast to the negligible unclassified proportion in the read data. For *Diaporthales* (contig data), one realm (*Duplodnaviria*) and one kingdom (*Heunggongvirae*) were detected at ~10% relative abundance (Fig. 9; Supplementary Table S22). Two samples (*Tubakia suttoniana* M24 and *Diaporthe vaccinii* M32) were particularly enriched in *Varidnaviria* (44–75%) (Supplementary Table S22). However, about 73% of viral contigs remained unclassified. In the read-based analysis, *Orthornavirae* dominated (76%), followed by *Heunggongvirae* (13%) and *Bamfordvirae* (11%). In *Glomerellales*, *Varidnaviria* was the most abundant resident realm in both datasets (35% in contigs, 62% in reads), followed by *Duplodnaviria* (23% in contigs, 29% in reads). Other realms were below 2% relative abundance. The most abundant kingdoms were *Bamfordvirae* (33% in contigs, 62% in reads)



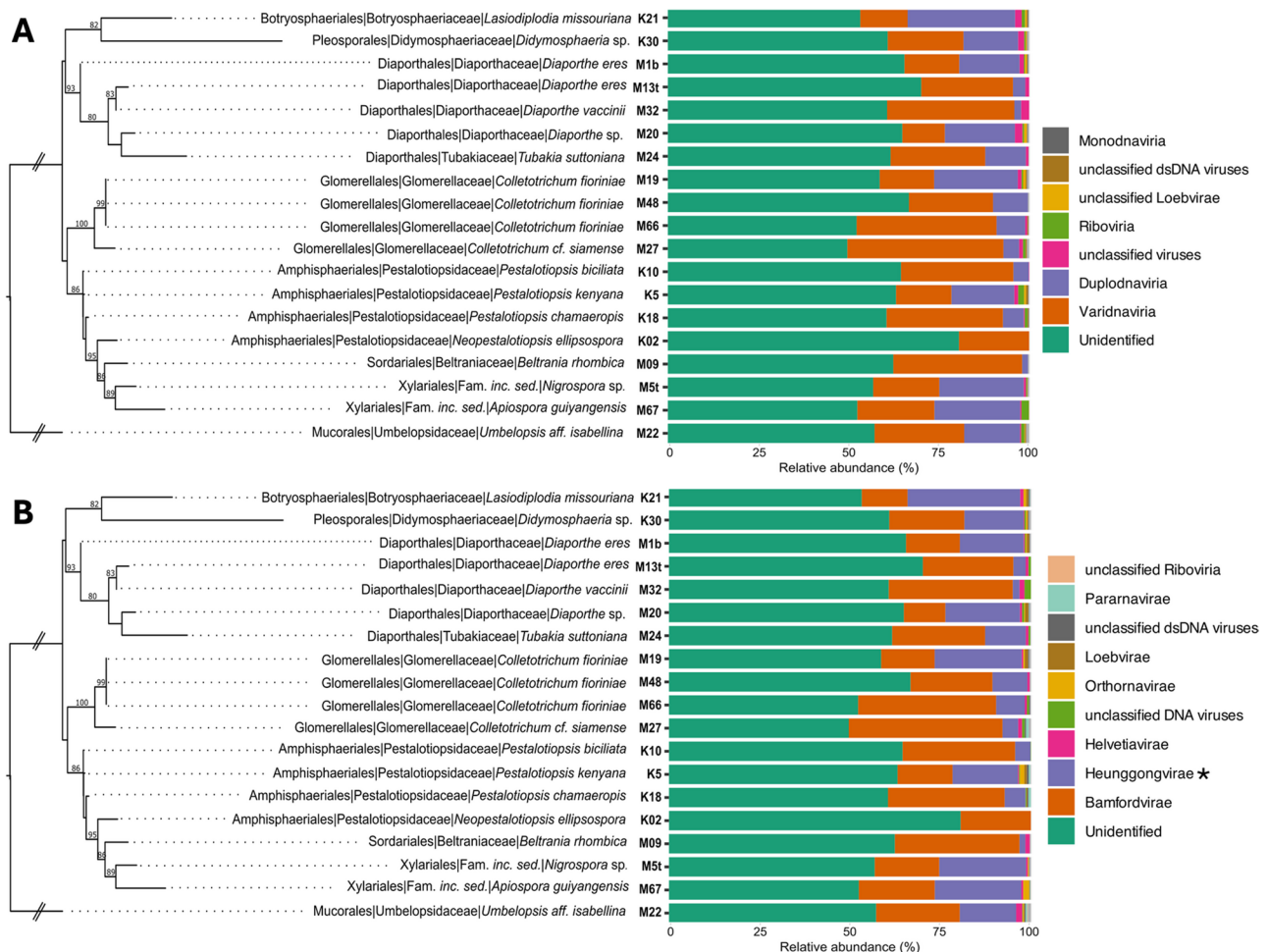
**Fig. 9** Relative abundance of the viral realms (A) and kingdoms (B) among the fungal isolates, using metatranscriptomic read data. The Maximum Likelihood phylogenetic tree was constructed using concatenated nrDNA ITS, TEF, and TUB sequences. Bootstrap values are shown on nodes. For B, potential taxa containing bacteriophages are indicated by \*

and *Heunggongvirae* (23% in contigs, 29% in reads), with other kingdoms contributing less than 2–8% (Figs. 9 and 10; Supplementary Tables S21 and S22). Around 39% of viral contigs in *Glomerellales* were unclassified at the realm or kingdom level, whereas all reads were successfully classified at these taxonomic levels.

Multilevel pattern and multinomial classification analyses revealed notable differences between the contig- and read-based approaches in detecting taxon-specific viral associations and phylogenetic signals (Table 1). Using the contig-based approach, multilevel pattern analysis identified significant viral associations primarily for *Glomerellales*. Taxa such as *Uroviricota* (*Duplodnaviria*) and *Nucleocytoviricota* (*Varidnaviria*) were significantly enriched in *Glomerellales* when compared to *Amphisphaeriales*, and *Naldaviricetes* (*Varidnaviria*) was significant in *Glomerellales* compared to *Diaporthales*

(Supplementary Figure S28). No significant taxa were identified for *Amphisphaeriales* or *Diaporthales*. The multinomial classification test detected some specialist taxa in *Diaporthales* compared to *Amphisphaeriales*; however, these belonged to unidentified viral groups. Comparisons between *Diaporthales* and *Glomerellales* indicated that specialist signals were mostly associated with *Megaviricetes* (*Varidnaviria*) (Supplementary Figures S29 and S30).

The read-based approach yielded more granular results in the multilevel pattern analysis. When comparing *Amphisphaeriales* to *Glomerellales*, 24 significant viral taxa were identified for *Amphisphaeriales* and 19 for *Glomerellales*, with *Megaviricetes* (*Varidnaviria*, *Bamfordvirae*) dominating in both groups (Supplementary Table S26). The comparison between *Amphisphaeriales* and *Diaporthales* revealed 10 significant taxa for the former



**Fig. 10** Relative abundance of the viral realms (A) and kingdoms (B) among the fungal isolates, using metatranscriptomic contig data. The Maximum Likelihood phylogenetic tree was constructed using concatenated nrDNA ITS, TEF, and TUB sequences. Bootstrap values are shown on nodes. For B, potential taxa containing bacteriophages are indicated by \*

(mostly *Megaviricetes*) and five for the latter, including two within *Duplodnaviria* (*Herviviricetes*, *Caudoviricetes*) and two within *Varidnaviria* (*Megaviricetes*, *Naldaviricetes*) (Supplementary Table S27). The *Diaporthales* vs. *Glomerellales* comparison revealed 18 significant taxa for *Diaporthales*—all belonging to *Caudoviricetes* (*Duplodnaviria*)—and two for *Glomerellales* (*Naldaviricetes* and *Caudoviricetes*) (Supplementary Table S28). In contrast to the multilevel pattern analysis, the multinomial test was found unsuitable due to the high number of “rare” taxa.

Phylogenetic signal analyses had mixed results (Table 1). With contig data, the Mantel test for the full dataset indicated no detectable phylogenetic signal, although a marginally significant result was observed in the balanced dataset containing the best-represented fungal orders. Both analyses yielded robust model fits after 10,000 permutations. The Procrustes analysis, however, showed a significant association between viral community composition and fungal host phylogeny

(Supplementary Figure S12C). The read-based approach also yielded no significant phylogenetic signal in the Mantel test for either the full dataset or the balanced subset, despite high model robustness. Yet, as with contigs, Procrustes analysis indicated a significant correlation between fungal hosts and viral community structure in both full and balanced datasets (Supplementary Figures S26E and S27E).

## Discussion

### Host specialization and community structure

Significant differences in beta diversity among fungal orders and families, combined with the presence of core bacterial and viral taxa, suggest a host-specific endohyphal community associated with the host fungus. This is further corroborated by the strong phylogenetic signal observed, particularly in bacterial communities, indicating that host association is influenced by phylogenetic relatedness. Such relatedness, in turn, plays a crucial role in shaping the composition of these microbial



communities [84, 85]. Notable bacterial taxa identified as indicator species or specialists for specific fungal groups include *Moraxellales*, *Sphingomonadales*, and *Streptosporangiaceae* for *Amphisphaeriales*; *Enterobacterales* (e.g., *Enterobacteraceae*), *Hyphomicrobiales* (e.g., *Rhizobiaceae*), and *Micrococcales* for *Glomerellales*; and *Cytophagales* for *Diaporthales*. While many additional taxa were classified as core, indicator and specialist bacterial taxa were encompassed within this core category.

In contrast, the mixed results for viral communities highlight the complexity of these interactions. Overall, there was no concordance between the core taxa classification, and the multilevel pattern (indicator) and multinomial (specialist) tests. Furthermore, although viral communities did not show significant differences across fungal orders and families in the full dataset, the significant phylogenetic signal in the balanced dataset suggests an underlying pattern of association when focusing on specific fungal orders. For example, some core, indicator, or specialist viral taxa may be closely associated with endohyphal bacteria or integrated within fungal genomes. Notably, *Caudoviricetes* (*Duplodnaviria*, *Uroviricota*, *Heunggongvirae*), which includes many bacteriophages, and *Megaviricetes* (*Varidnaviria*, *Nucleocytoviricota*, *Bamfordvirae*), found as possible endogenous elements in fungal genomes, illustrate these associations [49, 86]. There are also other examples of *Nucleocytoviricota* in protists, algae, and arthropods, among other eukaryotes [87, 88]. Therefore, the lack of significant differences in the full dataset; mixed results in core, indicator, and specialist taxa; but significant phylogenetic signal, suggest that other factors, such as the environment, mode of virus transmission, or molecular host-virus interactions [10, 41–43, 45, 46, 89–91], might also influence viral community composition and host specialization.

### Endohyphal bacterial communities

In the present work, a great diversity and abundance of endohyphal bacteria were found among the fungi sampled, supporting previous studies [11, 24, 92, 93]. Additionally, the lack of an asymptote in the taxa accumulation curves suggests that much greater diversity remains to be characterized. The detection of core bacterial taxa in the fungal samples suggests a level of specialization that could be vital for the functioning and stability of these symbiotic relationships [94–98]. Unraveling the functional roles of these bacteria in shaping fungal phenotypes, as well as their contributions to direct and indirect symbioses, is a critical area of research that warrants greater attention. This aligns with a growing body of research exploring how microbial partners modulate host phenotypes, including secondary metabolite production, pathogenicity, and stress tolerance in fungi, areas that are

increasingly recognized for their ecological and applied significance [20, 23, 24, 99–104]. Our results contribute to this broader effort by highlighting specific bacterial lineages that may influence fungal metabolism, dispersal, or interactions with plant hosts.

Some noteworthy bacteria that were found in our study belong to the *Bacillales*, *Burkholderiales*, *Enterobacterales*, *Hyphomicrobiales*, and *Pseudomonadales*, which are known to play a role in fungal phenotype and symbiosis establishment. For example, some endohyphal *Enterobacter* species can increase the expression of important toxins in fungi (e.g., fumonisin in *Fusarium fujikuroi*) [20] or change their plant pathogenicity by altering genes in their hosts [104]. Members of the *Burkholderiaceae*, which have been reported as abundant in *Mucoromycota* [105], play a significant role in modulating rhizoxin production, a macrocyclic polyketide metabolite and phytotoxin critical to plant pathogenicity [99, 100]. This compound, previously attributed to the pathogenic fungus *Rhizopus*, is actually produced by endohyphal *Burkholderia* species. *Burkholderiaceae*, in general, has consistently been found in many groups of fungi [97]. *Rhizobium radiobacter* (*Hyphomicrobiales*) is related to the successful endophytic colonization of the basidiomycete *Piriformospora indica* as well as to the improvement of the systemic protection of different crops against bacterial pathogens such as *Xantomonas traslucens* or *Pseudomonas syringae* [23, 24, 101]. Spores from *Glomerales* (arbuscular mycorrhizal fungi, AMF) associated with *Bacillus*, *Pseudomonas*, and *Rhizobium* bacteria suggest a recruited microbiota to increase eventual fungal dispersal and colonization success [103]. Additionally, evidence supports a tripartite, plant-mediated synergistic symbiosis among AMF, *Rhizobium*, and the plant, centered on nutrient exchange [102, 106]. While there is no well-established evidence of a direct cost-benefit or nutrient exchange between AMF and *Rhizobium*, some studies suggest potential signaling interactions or rhizosphere microbial community effects that may facilitate indirect mutual benefits [102]. Lastly, in our study, some bacterial orders detected through the metagenomic data have not been reported before as endohyphal (e.g., *Lactobacillales*, *Micrococcales*, *Moraxellales*, *Rhodospirillales*, and *Vibrionales*) and therefore, future studies could focus on elucidating the functional contributions of these newly identified bacterial taxa to fungal metabolism and fitness.

### Endohyphal viral communities

In contrast to the diversity of endohyphal bacteria, the viral sequence diversity in our study was lower. This is supported by taxa accumulation curves, which reached an asymptote for some fungal hosts. For instance, we only identified *Bamfordvirae* (*Varidnaviria*) and *Heunggongvirae* (*Duplodnaviria*), both with dsDNA genomes, as core

taxa across all fungal orders in the metagenomic data. No core taxa were identified in the metatranscriptomic data. These findings contradict other studies that reported *Riboviria* (i.e., dsRNA viruses) and *Monodnaviria* as the most abundant mycoviral realms [7, 10]. In our study, *Riboviria* sequences were detected in approximately 4% and 1% of the metagenomic and metatranscriptomic datasets, respectively, while *Monodnaviria* was absent in the metagenomic dataset and present at a negligible level (~0.04%) in the metatranscriptomic dataset. In addition, we expected to find more dsRNA mycoviruses, but we only detected a few contigs or reads of dsRNA mycoviruses (e.g., *Alternavirus*, *Diplodia scorbiculata* RNA virus 1-like, *Scleroulivirus*, and other unidentified dsRNA viruses). Despite these differences, roughly 40% of the viral contigs remained unclassified, highlighting the challenges in virome studies and the need for expanded and improved viral taxonomic databases [107–109].

The *Caudoviricetes* (*Duplodnaviria*, *Uroviricota*, *Heunggongvirae*), which are abundant across all fungal orders in our study and include bacteriophages, have the potential to influence fungal phenotypes and genomic architecture. Phages can indirectly affect fungi through interactions with bacterial intermediaries and horizontal gene transfer, playing a critical role in shaping fungal phenotypes and microbial ecosystems. For example, phages can transfer DNA between bacteria and, in some cases, to fungi, a process known as bacteriophage-mediated gene transfer, which, in consequence, can have significant ecological and evolutionary implications [110]. Bacteriophage-mediated gene transfer has been shown to influence fungal colonization of plant roots [110], fungal degradation processes [97], and the production of fungal secondary metabolites, such as terpenoids, which serve diverse ecological functions [111]. These metabolites can kill nematodes and arthropods, attract animals to facilitate spore dissemination, and mediate communication between fungi and bacteria [111]. Additionally, phages can influence fungi indirectly by affecting bacterial populations through bacterial lysis and community dynamics. Phage-mediated bacterial lysis releases key nutrients—such as nitrogen, phosphorus, and carbon—into the environment, promoting fungal growth and metabolism [112, 113]. In the rhizosphere, phages infecting nitrogen-fixing bacteria, such as *Rhizobium*, can disrupt nutrient cycling [114–116], thereby reducing resources available to fungi. Furthermore, phages can modulate the production of bacterial secondary metabolites, which often play a key role in bacterial-fungal interactions. For instance, phages infecting *Bacillus* species may enhance the production of antifungal compounds like iturin A, thereby altering fungal pathogen suppression [117, 118].

*Mimiviridae* and *Phycodnaviridae* (*Varidnaviria*, *Bamfordvirae*, *Megaviricetes*) were the most abundant

viral families. These families have been reported mostly in amoebae, algae, and bacteria. Although direct infections of fungi by members of *Nucleocytoviricota* (giant viruses; *Varidnaviria*, *Bamfordvirae*) are not well-documented [49, 86, 119], there is evidence of interactions through endogenous viral elements and environmental viromes. For example, a significant 1.5-Mb endogenous viral region, related to the family *Asfarviridae* within *Nucleocytoviricota*, was discovered in the genome of the arbuscular mycorrhizal fungus *Rhizophagus irregularis* [49]. This suggests ancient viral integration events that may influence fungal evolution and genome architecture. Studies have also identified viral sequences related to *Nucleocytoviricota* in various environments, including freshwater ecosystems [120]. These findings indicate the presence of giant viruses in habitats where fungi are prevalent, suggesting potential interactions. Additionally, research has uncovered complex genomes of early nucleocytoviruses through ancient endogenous viral elements in diverse eukaryotic lineages, including fungi [121]. This highlights the role of giant viruses in horizontal gene transfer and their potential impact on the evolution of fungal genomes.

#### Comparison of read- and contig-based approaches

Our comparative analysis of read- and contig-based approaches in metagenomic and metatranscriptomic datasets revealed shared patterns in microbial richness and community structure across fungal hosts, with both approaches detecting significant phylogenetic signals. However, notable differences emerged: the read-based approach was more sensitive in detecting microbial diversity and yielded stronger phylogenetic signal in bacterial communities, especially with the balanced dataset, while also identifying numerous bacterial indicator taxa. Yet, it struggled with taxonomic resolution, particularly for viruses, and rarely recovered core or resident taxa. In contrast, the contig-based approach detected fewer taxa overall but provided higher taxonomic resolution and more consistent recovery of core and resident taxa among bacteria and viral communities, making it more effective for identifying persistent symbionts. Its limitations included greater susceptibility to undersampling and lower statistical power in small datasets, particularly for viral phylogenetic signal.

Previous studies suggest that the choice of method for metagenomic microbial community profiling depends on the target environment and desired taxonomic resolution, with assembly-based approaches (particularly those using long reads) offering greater accuracy for species-level classification, while raw read analyses may be sufficient in well-characterized systems with comprehensive reference databases [122, 123]. Overall, in our study, the two approaches were complementary: read-based

analyses were better at capturing community-wide trends, high-resolution comparisons, and phylogenetic structuring, while contig-based analyses provided deeper taxonomic insights and more robust detection of ecologically persistent taxa.

### Limitations and future directions

The absence of visible bacterial-like cells in isolate M67, as observed under transmission electron microscopy (TEM), may indicate that the sectioned hyphae had not yet been colonized by endohyphal bacteria or that such bacteria were lost during subculturing [13]. In contrast, the clear presence of bacteria within the hyphae of isolates M5t and K21 provides direct support for the sequencing-based detection of endohyphal bacteria. However, virus-like particles were not detected in any of the isolates examined, suggesting that mycoviruses may occur at low titers—at least under the culture conditions used. This is consistent with the relatively low number of viral contigs and reads (when compared to bacteria) identified in the metagenomic and metatranscriptomic assemblies.

Studies show that several factors may contribute to the survival, apparent absence, or low abundance of endohyphal bacteria and viruses, including storage duration and temperature, repeated freeze-thaw cycles, subculturing, antibiotics, and in vitro culture conditions [13, 124, 125]. These factors can reduce the viability or persistence of both bacteria and viruses within fungal hosts over time, potentially leading to their underrepresentation or loss. In this study, we took measures to mitigate such effects and preserve the integrity of endohyphal communities. For example, antibiotics were used at low concentrations and only during the initial isolation step from leaf tissue; no antibiotics were used in subsequent culturing. In total, only two subculturing steps were performed before transferring the fungi to liquid culture for biomass production. DNA and RNA extractions were conducted from freshly harvested mycelium, and long-term storage at -80 °C occurred only after these extractions were completed. To improve the detection and study of endohyphal microbes, future efforts should continue to minimize these stressors and implement optimized protocols for fungal sample preservation. Additionally, targeted methods—such as concentrating virus particles from larger culture volumes or using PCR primers designed from assembled viral scaffolds—may further enhance the recovery and characterization of mycoviruses.

Another important limitation was the high percentage of unclassified taxa, which may affect interpretative power, underscoring the need for expanded reference databases, especially for viruses [107–109]. Notably, many of the specialist taxa (resulting from the multinomial classification) are unidentified. This

limitation reflects the underrepresentation of viral and bacterial sequences in reference databases and the complexity of bacterial and viral taxonomy [126–129]. This taxonomic gap is worsened by the diminishing number of trained taxonomists, a phenomenon known as the ‘taxonomic impediment,’ which hampers the discovery and classification of biodiversity [130].

Additional limitations or considerations should be acknowledged. Our sample size for certain fungal orders was also limited, which may have impacted the robustness of phylogenetic and community analyses. Some statistical outcomes, particularly phylogenetic signal in viral communities, were only significant in subset analyses. This variability likely reflects a combination of biological heterogeneity across fungal hosts and limitations in taxonomic resolution, especially for low-abundance or rare viral taxa [131]. We interpret these patterns as context-dependent rather than contradictory, and emphasize that such variability is expected in complex multipartite symbioses and should be considered when interpreting community-level trends [132, 133]. Expanding the sample size for underrepresented fungal orders and including additional orders will certainly strengthen inferences about the phylogenetic signal and host specialization, especially for endohyphal bacteria, which did not reach an asymptote in the species accumulation curve in our metagenomic data. We also did not consistently assess the presence of viral or bacterial sequences inserted into fungal genome assemblies (e.g., endogenous elements) or determine whether these insertions were recent or ancestral events [47–49]. Investigating viral and bacterial sequences integrated into fungal or eukaryotic genomes can provide compelling insights into evolutionary and functional impacts [47, 134, 135]. Lastly, the reliance on high-throughput sequencing may potentially introduce biases, such as uneven detection of RNA versus DNA viruses [136].

### Conclusions

This study is among the first to use phylogenetic signal analyses to explore host specialization in endohyphal microbial communities, offering insights into multipartite fungal-microbe interactions. By focusing on three fungal orders within *Sordariomycetes* (*Ascomycota*), we revealed ecological and phylogenetic associations with bacterial and viral communities. The presence of significant phylogenetic signals and consistent core bacterial taxa (e.g., *Bacillales*, *Burkholderiales*, *Enterobacterales*, *Hyphomicrobiales*, and *Pseudomonadales*) across fungal lineages supports the idea of specialized, potentially co-evolved relationships that may have implications for fungal phenotype and symbiosis establishment. In contrast, viral communities showed inconsistent phylogenetic structuring, with signals detected only in certain

datasets. The identification of *Caudoviricetes* and *Mega-viricetes* points to possible horizontal gene transfer and genomic integration events influencing bacterial and fungal phenotypes. These results highlight the complexity of plant-fungal-bacterial-viral interactions and the need for deeper functional and evolutionary investigations. These findings also reinforce the importance of microbial partners in potentially shaping fungal and plant biology, and ecosystem function. Continued efforts to explore these hidden microbial communities, expanding reference databases, and broader sampling efforts will be essential to fully understand the dynamics and consequences of these multipartite symbioses.

#### Abbreviations

ANOVA	Analysis of variance
BLAST	Basic Local Alignment Search Tool
BSA	Bovine serum albumin
CeNAT-CONARE	Centro Nacional de Alta Tecnología, Consejo Nacional de Rectores
dbRDA	Distance-based redundancy analysis
DMSO	Dimethyl sulfoxide
dsRNA/DNA	Double-stranded RNA/DNA
HPC	High-performance computing
ITS nrDNA	Internal transcribed spacer regions of the nuclear ribosomal DNA
MEB	Malt extract broth
NCBI RefSeq	National Center for Biotechnology Information, Reference Sequence Database
NMDS	Non-metric multidimensional scaling
PCR	Polymerase chain reaction
PDA	Potato dextrose agar
PERMANOVA	Permutational multivariate analysis of variance
TEF	Translation-elongation factor 1- $\alpha$
TEM	Transmission electron microscopy
TUB	$\beta$ -tubulin

#### Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s40793-025-00757-8>.

Supplementary Material 1

#### Acknowledgements

We would like to acknowledge Bowie State University students from the Course-based Undergraduate Research Experience (CURE) course Microbiology BIOL309 (Spring 2024) for their assistance in DNA and RNA extraction from four fungal samples.

#### Author contributions

Conception: PC, MC, JNC; design of the work: PC, JNC, EEL; data acquisition: PC, MB, NG, AD, JPK; data analysis: PC, EEL, MAS, NG, MB, JPK, JNC; data interpretation: PC, EEL, MAS, JK, JNC, MC; drafted the written work: EEL, PC; revised the written work: PC, EEL, MC, JNC, JPK, MB, NG, AD, MAS.

#### Funding

This study is funded by a U.S. National Science Foundation grant to P. Chaverri (IOS-2321265). Additional funding for the students' CURE project was provided by the Department of Natural Sciences (Bowie State University).

#### Data availability

Supplementary tables and figures supporting the results and conclusions of this article are available in the Zenodo repository (<https://doi.org/10.5281/zenodo.15620648>). Raw. fastq files for metagenomic and metatranscriptomic data

are deposited in GenBank's Sequence Read Archive (SRA) under BioProject PRJNA1221291. GenBank accession numbers for newly generated sequences (ITS nrDNA, TEF, and TUB) are indicated in Supplementary Table S1.

#### Declarations

##### Competing interests

The authors declare no competing interests.

##### Author details

<sup>1</sup>Centro De Investigaciones En Productos Naturales, Universidad De Costa Rica, San José, Costa Rica

<sup>2</sup>Centro Nacional De Innovaciones Biotecnológicas (CeNAT, CONARE), San José, Costa Rica

<sup>3</sup>Department of Natural Sciences, Bowie State University, Bowie, MD, USA

<sup>4</sup>Institute for Bioscience and Biotechnology Research, Department of Plant Science and Landscape Architecture, University of Maryland, College Park, MD, USA

<sup>5</sup>Centro De Investigación En Neurociencias, Universidad De Costa Rica, San José, Costa Rica

<sup>6</sup>Escuela De Biología, Universidad De Costa Rica, San José, Costa Rica

<sup>7</sup>Escuela De Química, Universidad De Costa Rica, San José, Costa Rica

Received: 27 March 2025 / Accepted: 16 July 2025

Published online: 25 July 2025

#### References

1. Dunn AM, Torchin ME, Hatcher MJ, Kotanen PM, Blumenthal DM, Byers JE, et al. Indirect effects of parasites in invasions. *Funct Ecol*. 2012;26:1262–74.
2. Strauss SY. Indirect effects in community ecology: their definition, study and importance. *Trends Ecol Evol*. 1991;6:206–10.
3. González-Teuber M, Palma-Onetto V, Aguilera-Sammaritano J, Mithöfer A. Roles of leaf functional traits in fungal endophyte colonization: potential implications for host–pathogen interactions. *J Ecol*. 2021;1–16.
4. Hubbard M, Germida JJ, Vujanovic V. Fungal endophytes enhance wheat heat and drought tolerance in terms of grain yield and second-generation seed viability. *J Appl Microbiol*. 2014;116:109–22.
5. Araldi-Brondolo SJ, Spraker J, Shaffer JP, Woytenko EH, Baltrus DA, Gallery RE, et al. Bacterial endosymbionts: master modulators of fungal phenotypes. Washington, DC, USA: ASM Press; Fungal Kingdom; 2017. pp. 981–1004.
6. Kelliher JM, Robinson AJ, Longley R, Johnson LYD, Hanson BT, Morales DP, et al. The endohyphal Microbiome: current progress and challenges for scaling down integrative multi-omic Microbiome research. *Microbiome*. 2023;11:1–14.
7. Myers JM, James TY. Mycoviruses *Curr Biol*. 2022;32:R150–5.
8. Pires MM, O'Donnell JL, Burkle LA, Díaz-Castelazo C, Hembry DH, Yeakel JD, et al. The indirect paths to cascading effects of extinctions in mutualistic networks. *Ecology*. 2020;101:1–8.
9. Montoya JM, Ecology. Dynamics of indirect extinction. *Curr Biol*. 2015;25:R1129–31.
10. Hough B, Steenkamp E, Wingfield B, Read D. Fungal viruses unveiled: A comprehensive review of mycoviruses. *Viruses*. 2023;15:1202.
11. Robinson AJ, House GL, Morales DP, Kelliher JM, Gallegos-Graves LV, LeBrun ES, et al. Widespread bacterial diversity within the bacteriome of fungi. *Commun Biol*. 2021;4:1168.
12. Deveau A, Bonito G, Uehling J, Paoletti M, Becker M, Bindschedler S, et al. Bacterial–fungal interactions: ecology, mechanisms and challenges. *FEMS Microbiol Rev*. 2018;42:335–52.
13. Hoffman MT, Arnold AE. Diverse bacteria inhabit living hyphae of phylogenetically diverse fungal endophytes. *Appl Environ Microbiol*. 2010;76:4063–75.
14. Shaffer JP, Sarmiento C, Zalamea P-C, Gallery RE, Davis AS, Baltrus DA, et al. Diversity, specificity, and phylogenetic relationships of endohyphal bacteria in fungi that inhabit tropical seeds and leaves. *Front Ecol Evol*. 2016;4:1–20.
15. Bastías DA, Johnson LJ, Card SD. Symbiotic bacteria of plant-associated fungi: friends or foes? *Curr Opin Plant Biol*. 2020;56:1–8.
16. Pakvaz S, Soltani J. Endohyphal bacteria from fungal endophytes of the mediterranean Cypress (*Cupressus sempervirens*) exhibit *in vitro* bioactivity. *Pathol*. 2016;46:569–81.



17. Pawlowska TE. Symbioses between fungi and bacteria: from mechanisms to impacts on biodiversity. *Curr Opin Microbiol.* 2024;80:102496.
18. Frey-Klett P, Burlinson P, Deveau A, Barret M, Tarkka M, Sarniguet A. Bacterial-fungal interactions: hyphens between agricultural, clinical, environmental, and food microbiologists. *Microbiol Mol Biol Rev.* 2011;75:583–609.
19. Obasa K, White FF, Fellers J, Kennelly M, Liu S, Katz B, et al. A dimorphic and virulence-enhancing endosymbiont bacterium discovered in *Rhizoctonia Solani*. *Phytobiomes J.* 2017;1:14–23.
20. Obasa K, Adesemoye A, Obasa R, Moraga-Amador D, Shinogle H, Alvarez S, et al. Endohyphal bacteria associated with virulence, increased expression of Fumonisin biosynthetic genes, and production of Fumonisin and macroconidia in *Fusarium Fujikuroi* W343. *Plant Pathol.* 2020;69:87–100.
21. Shaffer JP, U'Ren JM, Gallery RE, Baltus DA, Arnold AE. An endohyphal bacterium (*Chitinophaga, Bacteroidetes*) alters carbon source use by *Fusarium keratoplasticum* (*F. solani* species complex, Nectriaceae) *Front Microbiol.* 2017;8.
22. Richter I, Radosa S, Cseresnyés Z, Ferling I, Büttner H, Niehs SP, et al. Toxin-producing endosymbionts shield pathogenic fungus against micropredators. *mBio.* 2022;13:1–16.
23. Guo H, Glaeser SP, Alalid I, Imani J, Haghighi H, Kämpfer P, et al. The abundance of endofungal bacterium *Rhizobium radiobacter* (syn. *Agrobacterium tumefaciens*) increases in its fungal host *Piriformospora indica* during the tripartite Sebacinalean symbiosis with higher plants. *Front Microbiol.* 2017;8:1–13.
24. Sharma M, Schmid M, Rothballer M, Hause G, Zuccaro A, Imani J, et al. Detection and identification of bacteria intimately associated with fungi of the order *Sebacinales*. *Cell Microbiol.* 2008;10:2235–46.
25. Myers JM, Bonds AE, Clemons RA, Thapa NA, Simmons DR, Carter-House D, et al. Survey of early-diverging lineages of fungi reveals abundant and diverse mycoviruses. *mBio.* 2020;11:1–17.
26. Ghabrial S, Caston J, Jiang D, Nibert M, Suzuki N. 50-plus years of fungal viruses. *Virology.* 2015;479:356–68.
27. Ghabrial S, Suzuki N. Viruses of plant pathogenic fungi. *Annu Rev Phytopathol.* 2009;47:353–84.
28. Kondo H, Botella L, Suzuki N. Mycovirus diversity and evolution revealed/ inferred from recent studies. *Annu Rev Phytopathol.* 2022;60:307–36.
29. Villan Larios DC, Diaz Reyes BM, Pirovani CP, Loguerio LL, Santos VC, Góes-Neto A, et al. Exploring the mycovirus universe: identification, diversity, and biotechnological applications. *J Fungi.* 2023;9:361.
30. Espinal RBA, de Santana SF, Santos VC, Lizardo GNR, Silva RJS, Corrêa RX, et al. Uncovering a complex Virome associated with the Cacao pathogens *Ceratocystis Cacaofunesta* and *Ceratocystis fimbriata*. *Pathogens.* 2023;12:287.
31. Kanhayuwa L, Kotta-Loizou I, Özkan S, Gunning AP, Coutts RHA. A novel mycovirus from *Aspergillus fumigatus* contains four unique DsRNAs as its genome and is infectious as DsRNA. *Proc Natl Acad Sci U S A.* 2015;112:9100–5.
32. Kotta-Loizou I. Mycoviruses and their role in fungal pathogenesis. *Curr Opin Microbiol.* 2021;63:10–8.
33. Xu G, Zhang X, Liang X, Chen D, Xie C, Kang Z, et al. A novel hexa-segmented DsRNA mycovirus confers hypovirulence in the phytopathogenic fungus *Diaporthe pseudophoenicola*. *Environ Microbiol.* 2022;24:4274–84.
34. Zhang H, Xie J, Fu Y, Cheng J, Qu Z, Zhao Z, et al. A 2-kb mycovirus converts a pathogenic fungus into a beneficial endophyte for *Brassica* protection and yield enhancement. *Mol Plant.* 2020;13:1420–33.
35. Yu X, Li B, Fu Y, Jiang D, Ghabrial SA, Li G, et al. A geminivirus-related DNA mycovirus that confers hypovirulence to a plant pathogenic fungus. *Proc Natl Acad Sci.* 2010;107:8387–92.
36. Zhou L, Li X, Kotta-Loizou I, Dong K, Li S, Ni D, et al. A mycovirus modulates the endophytic and pathogenic traits of a plant associated fungus. *ISME J.* 2021;15:1893–906.
37. Zhu JZ, Guo J, Hu Z, Zhang XT, Li XG, Zhong J. A novel partitivirus that confer hypovirulence to the plant pathogenic fungus *Colletotrichum liriopes*. *Front Microbiol.* 2021;12:653809–653809.
38. Zhu JZ, Qiu ZL, Gao BD, Li XG, Zhong J. A novel partitivirus conferring hypovirulence by affecting vesicle transport in the fungus *Colletotrichum*. *mBio.* 2024;15:e02530–23.
39. Filippou C, Garrido-Jurado I, Meyling NV, Quesada-Moraga E, Coutts RHA, Kotta-Loizou I. Mycoviral population dynamics in Spanish isolates of the entomopathogenic fungus *Beauveria Bassiana*. *Viruses.* 2018;10:665.
40. Steffan BN, Venkatesh N, Keller NP. Let's get physical: Bacterial-fungal interactions and their consequences in agriculture and health. *J Fungi.* 2020;6:1–18.
41. Bastolla U, Fortuna MA, Pascual-García A, Ferrera A, Luque B, Bascompte J. The architecture of mutualistic networks minimizes competition and increases biodiversity. *Nature.* 2009;458:1018–20.
42. Lunau K. Adaptive radiation and coevolution - pollination biology case studies. *Org Divers Evol.* 2004;4:207–24.
43. Poisot T, Bever JD, Nemri A, Thrall PH, Hochberg ME. A conceptual framework for the evolution of ecological specialisation. *Ecol Lett.* 2011;14:841–51.
44. Yaegashi H, Nakamura H, Sawahata T, Sasaki A, Iwanami Y, Ito T, et al. Appearance of mycovirus-like double-stranded RNAs in the white root rot fungus, *Rosellinia necatrix*, in an Apple orchard. *FEMS Microbiol Ecol.* 2013;83:49–62.
45. García-Pedrajas MD, Cañizares MC, Sarmiento-Villamil JL, Jacquat AG, Dambolena JS. Mycoviruses in biological control: from basic research to field implementation. *Phytopathology.* 2019;109:1828–39.
46. Choi GH, Dawe AL, Churbanov A, Smith ML, Milgroom MG, Nuss DL. Molecular characterization of vegetative incompatibility genes that restrict hypovirus transmission in the chestnut blight fungus *Cryphonectria parasitica*. *Genetics.* 2012;190:113–27.
47. Gong Z, Zhang Y, Han G-Z. Molecular fossils reveal ancient associations of DsDNA viruses with several phyla of fungi. *Virus Evol.* 2020;6:veaa008.
48. Kondo H, Chiba S, Suzuki N. Detection and analysis of Non-retroviral RNA Virus-Like elements in plant, fungal, and insect genomes. In: Uyeda I, Masuta C, editors. *Plant virol Protoc.* New York, NY: Springer New York; 2015. pp. 73–88.
49. Zhao H, Zhang R, Wu J, Meng L, Okazaki Y, Hikida H, et al. A 1.5-Mb continuous endogenous viral region in the arbuscular mycorrhizal fungus *Rhizophagus irregularis*. *Virus Evol.* 2023;9:vead064.
50. Gazis R, Chaverri P. Diversity of fungal endophytes in leaves and stems of wild rubber trees (*Hevea brasiliensis*) in Peru. *Fungal Ecol.* 2010;3:240–54.
51. Escudero-Leyva E, Granados-Montero MDM, Orozco-Ortiz C, Araya-Valverde E, Alvarado-Picado E, Chaves-Fallas JM, et al. The endophytobiome of wild *Rubiaceae* as a source of antagonistic fungi against the American leaf spot of coffee (*Mycena citricolor*). *J Appl Microbiol.* 2023;134:ixad090.
52. Peel MC, Finlayson BL, McMahon TA. Updated world map of the Köppen-Geiger climate classification. *Hydrol Earth Syst Sci.* 2007;11:1633–44.
53. Greenfield M, Pareja R, Ortiz V, Gómez-Jiménez MI, Vega FE, Parsa S. A novel method to scale up fungal endophyte isolations. *Biocontrol Sci Technol.* 2015;25:1208–12.
54. Schoch CL, Seifert KA, Huhndorf S, Robert V, Spouge JL, Levesque CA, et al. Nuclear ribosomal internal transcribed spacer (ITS) region as a universal DNA barcode marker for *Fungi*. *Proc Natl Acad Sci.* 2012;109:6241–6.
55. Carbone I, Kohn LM. A method for designing primer sets for speciation studies in filamentous ascomycetes. *Mycologia.* 1999;91:553–6.
56. Glass NL, Donaldson GC. Development of primer sets designed for use with the PCR to amplify conserved genes from filamentous ascomycetes. *Appl Environ Microbiol.* 1995;61:1323–30.
57. Herrera CS, Rossman AY, Samuels GJ, Lechat C, Chaverri P. Revision of the genus *Corallomyces* with *Corallonectria* gen. nov. for *C. jatrophae* (Nectriaceae, Hypocreales). *Mycosytema.* 2013;32:518–44.
58. Sofi MY, Shafi A, Masoodi KZ. BioEdit in bioinformatics. *Bioinforma everyone.* Elsevier; 2022. pp. 231–6.
59. Gazis R, Rehner S, Chaverri P. Species delimitation in fungal endophyte diversity studies and its implications in ecological and biogeographic inferences. *Mol Ecol.* 2011;20:3001–13.
60. Edgar RC. MUSCLE: A multiple sequence alignment method with reduced time and space complexity. *BMC Bioinformatics.* 2004;5:1–19.
61. Tamura K, Stecher G, Kumar S. MEGA11: molecular evolutionary genetics analysis version 11. *Mol Biol Evol.* 2021;38:3022–7.
62. Kalyaanamoorthy S, Minh BQ, Wong TKF, Von Haeseler A, Jermiin LS. ModelFinder: fast model selection for accurate phylogenetic estimates. *Nat Methods.* 2017;14:587–9.
63. Minh BQ, Schmidt HA, Chernomor O, Schrempf D, Woodhams MD, Von Haeseler A, et al. IQ-TREE 2: new models and efficient methods for phylogenetic inference in the genomic era. *Mol Biol Evol.* 2020;37:1530–4.
64. Nurk S, Meleshko D, Korobeynikov A, Pevzner PA. MetaSPAdes: A new versatile metagenomic assembler. *Genome Res.* 2017;27:824–34.
65. Bushmanova E, Antipov D, Lapidus A, Pribelski AD. RnaSPAdes: A de Novo transcriptome assembler and its application to RNA-Seq data. *GigaScience.* 2019;8:1–13.
66. Menzel P, Ng KL, Krogh A. Fast and sensitive taxonomic classification for metagenomics with Kaiju. *Nat Commun.* 2016;7.
67. Ye SH, Siddle KJ, Park DJ, Sabeti PC. Benchmarking metagenomics tools for taxonomic classification. *Cell.* 2019;178:779–94.

68. Buffet-Bataillon S, Rizk G, Cattoir V, Sassi M, Thibault V, Del Giudice J, et al. Efficient and quality-optimized metagenomic pipeline designed for taxonomic classification in routine Microbiological clinical tests. *Microorganisms*. 2022;10:711.
69. Faoro F, Faccio A, Balestrini R. Contributions of ultrastructural studies to the knowledge of filamentous fungi biology and fungi-plant interactions. *Front Fungal Biol*. 2022;2:805739.
70. Spurr AR. A low-viscosity epoxy resin embedding medium for electron microscopy. *J Ultrastruct Res*. 1969;26:31–43.
71. Reynolds ES. The use of lead citrate at high pH as an electron-opaque stain in electron microscopy. *J Cell Biol*. 1963;17:208–12.
72. R Core Team. R: A language and environment for statistical computing. Vienna, Austria; 2020.
73. Hsieh TC, Ma KH, Chao A. iNEXT: an R package for rarefaction and extrapolation of species diversity (Hill numbers). *Methods Ecol Evol*. 2016;7:1451–6.
74. Oksanen AJ, Blanchet FG, Friendly M, Kindt R, Legendre P, Mcglinn D et al. Package 'vegan'. 2012.
75. De Cáceres M, Legendre P, Wiser SK, Brotons L. Using species combinations in indicator value analyses. *Methods Ecol Evol*. 2012;3:973–82.
76. Chazdon RL, Chao A, Colwell RK, Lin S, Norden N, Letcher G, et al. A novel statistical method for classifying habitat generalists and specialists. *Ecology*. 2015;92:1332–43.
77. Ainsworth D, Krause T, Bridge L, Torda T, Raina G, Zakrzewski J-B. The coral core Microbiome identifies rare bacterial taxa as ubiquitous endosymbionts. *ISME J*. 2015;9:2261–74.
78. Hernandez-Agreda A, Leggat W, Bongaerts P, Herrera C, Ainsworth TD. Rethinking the coral microbiome: simplicity exists within a diverse microbial biosphere. *mBio*. 2018;9:e00812–18.
79. Vestergaard SZ, Dottorini G, Peces M, Murguz A, Dueholm MKD, Nierychlo M, et al. Microbial core communities in activated sludge plants are strongly affected by immigration and geography. *Environ Microbiome*. 2024;19:63.
80. Zhang X, Zhao W, Kou Y, Fang K, Liu Y, He H, et al. The contrasting responses of abundant and rare microbial community structures and co-occurrence networks to secondary forest succession in the subalpine region. *Front Microbiol*. 2023;14:1177239.
81. Liu C, Cui Y, Li X, Yao M. *microeco*: an R package for data mining in microbial community ecology. *FEMS Microbiol Ecol*. 2021;97:faa255.
82. Venables WN, Ripley BD. Modern applied statistics with S. 4th ed. New York: Springer; 2002.
83. Peres-Neto PR, Jackson DA. How well do multivariate data sets match? The advantages of a procrustes superimposition approach over the mantel test. *Oecologia*. 2001;129:169–78.
84. Näpflin K, Schmid-Hempel P. Host effects on microbiota community assembly. *J Anim Ecol*. 2018;87:331–40.
85. Schwob G, Almendras K, Veas-Mattheos K, Pezoa M, Orlando J. Host specialization and spatial divergence of bacteria associated with *Peltigera* lichens promote landscape gamma diversity. *Environ Microbiome*. 2024;19:57.
86. Bhattacharjee AS, Schulz F, Woyke T, Orcutt BN, Martínez JM. Genomics discovery of giant fungal viruses from subsurface oceanic crustal fluids. *ISME Commun*. 2023;3:10.
87. Scola BL, Audic S, Robert C, Jungang L, De Lamballerie X, Drancourt M, et al. A giant virus in amoebae. *Science*. 2003;299:2033–2033.
88. Woo AC, Gaia M, Guglielmini J, Da Cunha V, Forterre P. Phylogeny of the *Varidnaviria* morphogenesis module: congruence and incongruence with the tree of life and viral taxonomy. *Front Microbiol*. 2021;12:704052.
89. Mojica KDA, Brussaard CPD. Factors affecting virus dynamics and microbial host-virus interactions in marine environments. *FEMS Microbiol Ecol*. 2014;89:495–515.
90. Rothenburg S, Brennan G. Species-specific host-virus interactions: implications for viral host range and virulence. *Trends Microbiol*. 2020;28:46–56.
91. Yaegashi H, Yoshikawa N, Ito T, Kanematsu S. A mycoreovirus suppresses RNA Silencing in the white root rot fungus, *Rosellinia necatrix*. *Virology*. 2013;444:409–16.
92. Bertaux J, Schmid M, Hutzler P, Hartmann A, Garbaye J, Frey-Klett P. Occurrence and distribution of endobacteria in the plant-associated mycelium of the ectomycorrhizal fungus *Laccaria bicolor* S238N. *Environ Microbiol*. 2005;7:1786–95.
93. Bonfante P, Desirò A. Who lives in a fungus? The diversity, origins and functions of fungal endobacteria living in *Mucoromycota*. *ISME J*. 2017;11:1727–35.
94. Giger GH, Ernst C, Richter I, Gassler T, Field CM, Sintsova A, et al. Inducing novel endosymbioses by implanting bacteria in fungi. *Nature*. 2024;635:415–22.
95. Hestrin R, Kan M, Lafer M, Wollard J, Kimbrel JA, Ray P, et al. Plant-associated fungi support bacterial resilience following water limitation. *ISME J*. 2022;16:2752–62.
96. Krespach MKC, Stroe MC, Netzker T, Rosin M, Zehner LM, Komor AJ, et al. *Streptomyces* polyketides mediate bacteria–fungi interactions across soil environments. *Nat Microbiol*. 2023;8:1348–61.
97. Pratama AA, van Elsas JD. Gene mobility in microbiomes of the mycosphere and mycorrhizosphere—role of plasmids and bacteriophages. *FEMS Microbiol Ecol*. 2019;95:fiz053.
98. Seipke RF, Kaltenpoth M, Hutchings MI. *Streptomyces* as symbionts: an emerging and widespread theme? *FEMS Microbiol Rev*. 2012;36:862–76.
99. Araldi-Brondolo SJ, Spraker J, Shaffer JP, Woytenko EH, Baltrus DA, Gallery RE, et al. Bacterial endosymbionts: master modulators of fungal phenotypes. In: Heitman J, Howlett BJ, Crous PW, Stukenbrock EH, James TY, Gow NAR, editors. *Fungal Kingd*. Washington, DC, USA: ASM; 2017. pp. 981–1004.
100. Partida-Martinez LP, Hertweck C. Pathogenic fungus harbours endosymbiotic bacteria for toxin production. *Nature*. 2005;437:884–8.
101. Waller F, Achatz B, Baltruschat H, Fodor J, Becker K, Fischer M, et al. The endophytic fungus *Piriformospora indica* reprograms barley to salt-stress tolerance, disease resistance, and higher yield. *Proc Natl Acad Sci*. 2005;102:13386–91.
102. Calderon RB, Dangi SR. Arbuscular mycorrhizal fungi and *Rhizobium* improve nutrient uptake and microbial diversity relative to dryland site-specific soil conditions. *Microorganisms*. 2024;12:667.
103. Agnolucci M, Battini F, Cristani C, Giovannetti M. Diverse bacterial communities are recruited on spores of different arbuscular mycorrhizal fungal isolates. *Biol Fertil Soils*. 2015;51:379–89.
104. Zhang P, Huguet-Tapia J, Peng Z, Liu S, Obasa K, Block AK, et al. Genome analysis and hyphal movement characterization of the hitchhiker endophyte *Enterobacter* sp. from *Rhizoctonia Solani*. *Appl Environ Microbiol*. 2024;90:1–22.
105. Okrasinski A, Bokus A, Duk K, Gęsiorska A, Sokołowska B, Miłobędzka A, et al. New endophyte relationships between *Mucoromycota* and *Burkholderiaceae* representatives. *Appl Environ Microbiol*. 2021;87:e02707–20.
106. Kasanek SA, Cheeke TE, Moran JJ, Roley SS. Tripartite interactions among free-living, N-fixing bacteria, arbuscular mycorrhizal fungi, and plants: mutualistic benefits and community response to co-inoculation. *Soil Sci Soc Am J*. 2024;88:1000–13.
107. Chang W-S, Harvey E, Mahar JE, Firth C, Shi M, Simon-Loriere E, et al. Improving the reporting of metagenomic virome-scale data. *Commun Biol*. 2024;7:1687.
108. Cobbin JC, Charon J, Harvey E, Holmes EC, Mahar JE. Current challenges to virus discovery by meta-transcriptomics. *Curr Opin Virol*. 2021;51:48–55.
109. Smits SL, Bodewes R, Ruiz-Gonzalez A, Baumgärtner W, Koopmans MP, Osterhaus ADME, et al. Assembly of viral genomes from metagenomes. *Front Microbiol*. 2014;5:714.
110. Bruto M, Prigent-Combaret C, Luis P, Moënne-Loccoz Y, Muller D. Frequent, independent transfers of a catabolic gene from bacteria to contrasted filamentous eukaryotes. *Proc R Soc B Biol Sci*. 2014;281:20140848.
111. Jia Q, Chen X, Köllner TG, Rinkel J, Fu J, Labbé J, et al. Terpene synthase genes originated from bacteria through horizontal gene transfer contribute to terpenoid diversity in fungi. *Sci Rep*. 2019;9:9223.
112. Middelboe M, Jørgensen NOG. Viral lysis of bacteria: an important source of dissolved amino acids and cell wall compounds. *J Mar Biol Assoc U K*. 2006;86:605–12.
113. Yang K, Wang X, Hou R, Lu C, Fan Z, Li J, et al. Rhizosphere phage communities drive soil suppressiveness to bacterial wilt disease. *Microbiome*. 2023;11:16.
114. Braga LPP, Spor A, Kot W, Breuil M-C, Hansen LH, Setubal JC, et al. Impact of phages on soil bacterial communities and nitrogen availability under different assembly scenarios. *Microbiome*. 2020;8:52.
115. Buée M, De Boer W, Martin F, Van Overbeek L, Jurkevitch E. The rhizosphere zoo: an overview of plant-associated communities of microorganisms, including phages, bacteria, archaea, and fungi, and of some of their structuring factors. *Plant Soil*. 2009;321:189–212.
116. Wang X, Tang Y, Yue X, Wang S, Yang K, Xu Y, et al. The role of rhizosphere phages in soil health. *FEMS Microbiol Ecol*. 2024;100:fae052.
117. Lv Z, Li R, Shi C, Lu Z, Meng F, Bie X. The efficient synthesis strategy of bacillomycin D in *Bacillus amyloliquefaciens* FmbJ. *Process Biochem*. 2024;145:131–8.

118. Yaraguppi DA, Bagewadi ZK, Patil NR, Mantri N. Iturin: A promising Cyclic lipopeptide with diverse applications. *Biomolecules*. 2023;13:1515.
119. Rodrigues RAL, Dos Santos Silva LK, Dornas FP, De Oliveira DB, Magalhães TFF, Santos DA, et al. Mimivirus fibrils are important for viral attachment to the microbial world by a diverse glycoside interaction repertoire. *J Virol*. 2015;89:11812–9.
120. Elbehery AHA, Deng L. Insights into the global freshwater Virome. *Front Microbiol*. 2022;13:953500.
121. Kijima S, Hikida H, Delmont TO, Gaia M, Ogata H. Complex genomes of early nucleocytoviruses revealed by ancient origins of viral aminoacyl-tRNA synthetases. *Mol Biol Evol*. 2024;41:msae149.
122. Tamames J, Cobo-Simón M, Puente-Sánchez F. Assessing the performance of different approaches for functional and taxonomic annotation of metagenomes. *BMC Genomics*. 2019;20:960.
123. Tran Q, Phan V. Assembling reads improves taxonomic classification of species. *Genes*. 2020;11:946.
124. Gould EA. Methods for long-term virus preservation. *Mol Biotechnol*. 1999;13:57–66.
125. Khan HA, Baig DI, Bhatti MF. An overview of mycoviral curing strategies used in evaluating fungal host fitness. *Mol Biotechnol*. 2023;65:1547–64.
126. Pallen MJ, Telatin A, Oren A. The next million names for Archaea and Bacteria. *Trends Microbiol*. 2021;29:289–98.
127. Caetano-Anollés G, Claverie J-M, Nasir A. A critical analysis of the current state of virus taxonomy. *Front Microbiol*. 2023;14:1240993.
128. Harris HMB, Hill C. A place for viruses on the tree of life. *Front Microbiol*. 2021;11:604048.
129. Bobay L-M. The prokaryotic species concept and challenges. In: Tettelin H, Medini D, editors. *The pangenome*. Cham: Springer International Publishing; 2020. pp. 21–49.
130. Engel MS, Ceriaco LMP, Daniel GM, Dellapé PM, Löbl I, Marinov M, et al. The taxonomic impediment: a shortage of taxonomists, not the lack of technical approaches. *Zool J Linn Soc*. 2021;193:381–7.
131. Neu AT, Allen EE, Roy K. Defining and quantifying the core microbiome: challenges and prospects. *Proc Natl Acad Sci*. 2021;118:e2104429118.
132. Yan H, Lin D, Gu G, Huang Y, Hu X, Yu Z, et al. Taxonomic dependency and Spatial heterogeneity in assembly mechanisms of bacteria across complex coastal waters. *Ecol Process*. 2024;13:6.
133. Donald ML, Bohner TF, Kolis KM, Shadow RA, Rudgers JA, Miller TEX. Context-dependent variability in the population prevalence and individual fitness effects of plant–fungal symbiosis. Wilson G, editor. *J Ecol*. 2021;109:847–59.
134. Feschotte C, Gilbert C. Endogenous viruses: insights into viral evolution and impact on host biology. *Nat Rev Genet*. 2012;13:283–96.
135. Leclercq S, Théze J, Chebbi MA, Giraud I, Moumen B, Ernenwein L, et al. Birth of a W sex chromosome by horizontal transfer of *Wolbachia* bacterial symbiont genome. *Proc Natl Acad Sci*. 2016;113:15036–41.
136. Fitzpatrick AH, Rupnik A, O'Shea H, Crispie F, Keaveney S, Cotter P. High throughput sequencing for the detection and characterization of RNA viruses. *Front Microbiol*. 2021;12:621719.

## Publisher's note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.