

PREPRINT

Author-formatted, not peer-reviewed document posted on 26/05/2026

DOI: <https://doi.org/10.3897/arphapreprints.e201349>

Combined strategies to evaluate fungal rhizosphere communities using Nanopore sequencing

 **Claudia Barrera Garzon, Karin Pritsch, Fabian Weigl**

1 Combined strategies to evaluate fungal rhizosphere communities 2 using Nanopore sequencing

3
4 Claudia Barrera^{1*}: claudia.barrera@tum.de, ORCID: 0000-0001-8936-5441

5 Karin Pritsch^{1,2}: karin.pritsch@tum.de, ORCID: 0000-0001-6384-2473

6 Fabian Weigl¹: fabian.weigl@tum.de, ORCID: 0000-0003-3973-6341

7
8 ¹ Ecophysiology of Plants, Professorship for Land Surface-Atmosphere Interactions, Research
9 Department Life Science Systems, School of Life Sciences, Technical University of Munich,
10 Freising, Germany

11 ² Research Unit Environmental Simulation, Helmholtz Center Munich, Ingolstaedter
12 Landstrasse 1, 85764 Neuherberg, Germany

13 14 *Corresponding author

15 Claudia Barrera, Ecophysiology of Plants, Professorship for Land Surface-Atmosphere
16 Interactions, Research Department Life Science Systems, School of Life Sciences, Technical
17 University of Munich, Germany.

18 Email: claudia.barrera@tum.de

19 20 Abstract

21 Evaluating the diversity of soil fungal communities is critical to understanding their role in
22 plant adaptation and ecology. Improving the resolution of fungal identification using new
23 sequencing techniques, such as Oxford Nanopore, which yield long read lengths, is therefore
24 highly desirable. Yet, the implementation of this sequencing technology for fungi and other
25 eukaryotes has just begun.

26 In this study, we evaluated different primer combinations that cover the entire internal
27 transcribed spacer (ITS) region and include fragments of the small (SSU) and large subunit
28 (LSU) regions. We designed a pipeline to recover most of the fungal diversity, compared two
29 classifiers and two databases, and added a peptide nucleic acid (PNA) to inhibit the co-
30 amplification of plant ITS.

31 Our results showed that including the LSU did not improve the resolution of individual-strain
32 classification, and that non-degenerate forward primers were more effective for ITS sequence
33 extraction, recovering up to 91% of strains in a mock community with an accuracy of 0.97.
34 Moreover, results with rhizosphere samples showed that adding PNA effectively removed host
35 contamination across most primer combinations, thereby improving read retention during data
36 processing. Additionally, host decontamination and ITS extraction steps enhanced the recovery
37 of fungal groups, mainly in the Ascomycota phylum.

38 Altogether, our study provides strategies for handling samples with high plant tissue
39 concentrations, from primer selection to fungal species classification, and presents a modular
40 pipeline to facilitate data processing tailored to the user's needs.

42 **Keywords:** classifier, full-length ITS, fungal diversity, mock community, pipeline, primer
43 selection.
44

45 **Introduction**

46 Fungi are ubiquitous and highly diverse organisms that play crucial roles in various
47 environments. In soil, they facilitate the decomposition of organic matter, regulate nutrient
48 cycling, enhance nutrient uptake in the plant-soil interface, and interact with other soil
49 microorganisms, thereby enhancing ecosystem complexity and stability (Went and Stark 1968;
50 De Menezes et al. 2017). Moreover, fungi interact with plants as phytopathogens, saprophytes,
51 or symbionts such as mycorrhizae or endophytes. These symbiotic interactions enhance soil
52 and rhizosphere health, stimulate plant growth and productivity, and mitigate biotic and abiotic
53 stresses (Itoo and Reshi 2013; Bender et al. 2014; Zeilinger et al. 2015; Akter et al. 2025).
54 Consequently, robust tools for characterizing fungal communities at the root-soil interface are
55 essential for understanding and preserving soil and rhizosphere ecosystems.

56 For decades, Illumina has become the standard approach for metabarcoding analyses utilizing
57 fragments of up to 600 bp (2×300 bp). The internal transcribed spacer (ITS) ribosomal DNA
58 has become the accepted barcode for fungal identification (Blaalid et al. 2013). Located
59 between the small subunit (SSU) and large subunit (LSU) rRNA genes, the ITS region consists
60 of ITS1 and ITS2 separated by the 5.8S rRNA gene, and varies in length from 300 to 1200 bp
61 (Heeger et al. 2019). The accumulated variation within the non-coding ITS, combined with
62 conserved flanking sites, renders this marker a suitable choice for discriminating between
63 fungal groups (Schoch et al. 2012). However, the short-read strategy of Illumina constrains the
64 target to either ITS1 or ITS2 for fungal identification.

65 Although ITS1 and ITS2 individually provide broad taxonomic coverage, the complex
66 evolutionary histories of certain fungal groups necessitate the use of multiple loci for accurate
67 classification (Lücking et al. 2020). In particular, the ITS region of some Ascomycetes,
68 including important saprophytes and phytopathogens, lacks sufficient resolution for species-
69 level discrimination (Lücking et al. 2020; Tedersoo et al. 2022). Therefore, the inclusion of the
70 SSU or LSU can favor the taxonomic classification (Nilsson et al. 2018). Third-generation
71 sequencing platforms, such as Pacific Biosciences (PacBio) and Oxford Nanopore
72 Technologies (ONT), address this limitation by enabling the sequencing of long amplicons
73 spanning several kilobases (Goodwin et al. 2016). However, achieving high-quality results
74 requires careful primer selection, optimization of long-amplicon PCR conditions, and
75 appropriate data-processing workflows (Tedersoo et al. 2022).

76 While long-read platforms have been widely adopted for bacterial metabarcoding using full-
77 length 16S rRNA genes (~1,500 bp) (Zhang et al. 2023; Buetas et al. 2024; Butler et al. 2025),
78 their application to fungal communities remains a topic of debate. Recent studies using PacBio
79 have successfully characterized fungal diversity using full ITS regions coupled with tailored
80 bioinformatic pipelines (Heeger et al. 2018; Tedersoo and Anslan 2019; Tedersoo et al. 2021).
81 Nevertheless, the still limited distribution of the PacBio platform and its moderate cost-
82 effectiveness might reduce its appeal among the scientific community (Cuber et al. 2023).
83 Hence, ONT arises as an affordable and more flexible solution. However, differences in the
84 preparation of libraries, raw data, and error profiles between PacBio and ONT necessitate the
85 development of a Nanopore-specific workflow for fungal metabarcoding.

86 Several studies have explored optimal target regions for Nanopore-based fungal amplicon
87 sequencing. Ohta et al. (2023) evaluated fragments spanning from the SSU to the LSU using
88 various reverse primers flanking ITS1, ITS2, and multiple LSU regions. Although they

89 observed no substantial improvement in taxonomic resolution beyond the D3 region of the LSU
 90 using the LR5 primer (Vilgalys & Hester, 1990), the lack of a database that systematically
 91 collects the complete operon limited the interpretation of read-length effects (Porrás-Alfaro et
 92 al. 2014). To address this, Lu et al. (2023) compiled a database containing sequences from the
 93 SSU to the LSU (~5.5 kb), although many taxa remained unresolved due to the scarcity of full-
 94 operon sequences. More recently, the Eukaryome database was developed as a community-
 95 curated reference for the complete rRNA operon across eukaryotes (Tedersoo et al. 2024),
 96 although its applicability for Nanopore sequencing has not yet been explored.

97 The analysis of environmental samples such as soil, rhizosphere, and roots presents additional
 98 challenges, particularly due to the presence of host plant DNA. This issue is especially
 99 pronounced in studies of intricate symbioses such as ectomycorrhizae. The use of fungal-
 100 specific primers, including ITS1catta (Tedersoo & Anslan, 2019) and LR5-F (Tedersoo et al.
 101 2008), could reduce plant co-amplification (Nilsson et al. 2018). Furthermore, peptide nucleic
 102 acids (PNA), a synthetic oligomer that binds to target sequences and blocks their amplification,
 103 have proven effective in suppressing plant ITS amplification in short-read sequencing
 104 (Lundberg et al. 2013; Cregger et al. 2018; Whitaker 2025). However, their performance in
 105 long-amplicon sequencing has not yet been validated. Alternatively, host-derived reads can be
 106 filtered during bioinformatic processing.

107 Although ONT provides the EPI2ME platform for genomic analyses, its functionality is limited
 108 for fungal metabarcoding. For example, it lacks tools such as ITSx for ITS extraction
 109 (Bengtsson-Palme et al. 2013) and vsearch for chimera detection and clustering (Rognes et al.
 110 2016), among others, which have been incorporated into customized pipelines for long-read
 111 fungal data (Rognes et al. 2016; Heeger et al. 2018; Özkurt et al. 2022). For prokaryotes,
 112 algorithms such as Emu have shown promising results in classifying full-length 16S rRNA from
 113 Nanopore sequencing (Curry et al. 2022). Despite these advances, a unified and optimized
 114 pipeline for processing fungal Nanopore long-read data remains lacking.

115 Considering the above, this study aims to establish an optimal combination of strategies for
 116 fungal community assessment using Nanopore sequencing, from primer selection to data
 117 processing. Fifteen primer combinations were evaluated, including those previously applied in
 118 PacBio-based studies (Tedersoo et al. 2018; Tedersoo and Anslan 2019), and extended targets
 119 spanning the ITS region into the LSU rRNA gene (Ohta et al. 2023) to compare taxonomic
 120 resolution across fragment lengths. A tailored bioinformatic pipeline was developed,
 121 incorporating quality filtering, chimera removal, host decontamination, and extraction of the
 122 complete ITS region. Taxonomic classification was assessed using two classifiers (EMU and
 123 SINTAX implemented in vsearch) and two reference databases (UNITE and Eukaryome). Due
 124 to the lack of a commercial mock fungal community, a custom mock community was assembled
 125 using soil-associated fungi and additional reference strains to evaluate methodological
 126 accuracy. The effectiveness of a PNA clamp in reducing plant ITS amplification was also tested.
 127 Finally, the proposed strategies were validated using mixed fine-root and rhizosphere samples
 128 from ectomycorrhiza-associated trees, and compared based on criteria such as read retention,
 129 taxonomic resolution, and richness of their major root-associated lineages (Ascomycota and
 130 Basidiomycota).

131

132

133

134

135 **Materials and Methods**

136 **DNA extraction, library preparation, and sequencing**

137 *DNA extraction of individual fungal strains and mock community preparation*

138 A fungal mock community was prepared using 26 fungal strains to allow the comparison of the
 139 different methods. The strains were selected to represent a mixture of fungi from the most
 140 abundant phyla found in environmental samples, including those from deciduous forest soils.
 141 Fungal DNA was isolated from pure cultures of strains provided by the Group of Fungal
 142 Biotechnology in Wood Science (TUM School of Life Sciences, Technical University of
 143 Munich, Freising, Germany), and from fruiting bodies collected in the field (Table S1). The
 144 cultured strains were grown on potato dextrose agar (PDA), and the mycelia were scraped from
 145 the medium surface, resuspended in PBS supplemented with 1% DMSO, and stored at -20°C
 146 until DNA extraction. 2 g of the fruiting bodies were ground in liquid nitrogen using sterilized
 147 pestles and mortars. Then, the resuspended mycelia and the ground tissue were used as initial
 148 material for DNA extraction using the DNeasy PowerSoil Pro Kit (Qiagen, Germany) according
 149 to the manufacturer's instructions. The quality of the extracted DNA was evaluated in 1%
 150 Agarose gel electrophoresis, and the quantities were measured using the Quantifluor dsDNA
 151 system (Promega, Madison, WI, USA). Then, the DNA from each specimen was diluted to a
 152 concentration between 0.1 and 0.5 ng.

153

154 *DNA extraction of rhizosphere samples*

155 Eight soil drill cores (diameter: 1.5 cm, depth: 25 cm) were taken at 1 – 1.5 m from the stem of
 156 three European beech trees (*Fagus sylvatica* L.) in 2023 in Freising, Germany (11°39'17"E,
 157 48°24'25"N, 508 m a.s.l). After pooling the cores per tree, the samples were transported on ice
 158 to the lab and processed the same day. The fine roots with attached soil that could not be
 159 manually shaken off were sorted out using sterilized tweezers and ground in liquid nitrogen
 160 using sterilized pestles and mortars. Around 250 mg of ground material was submitted to DNA
 161 extraction using the DNeasy PowerSoil Pro Kit. The quality of the extracted material was
 162 evaluated in 1% Agarose gel electrophoresis, and the quantity was measured using the
 163 Quantifluor dsDNA system. Additionally, the DNA of two samples collected and processed in
 164 the same way in 2020 in Bischofswiesen, Germany (12°55'09.9"E, 47°39'59.2"N, 764 m a.s.l.)
 165 in a beech forest were used as a reference to compare with sequencing results obtained from
 166 the same samples using Illumina (Heym et al. 2023).

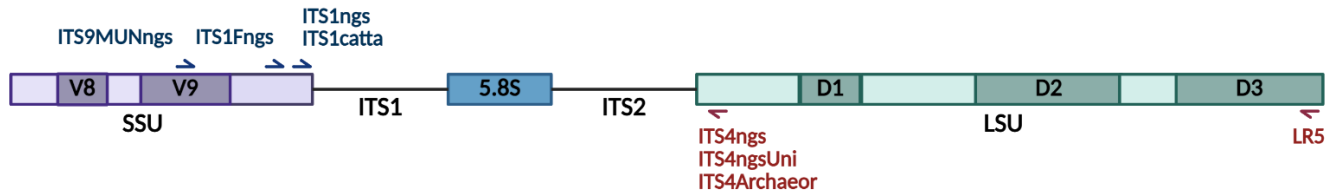
167

168 *ITS amplification*

169 The complete ITS or the ITS + LSU (Fig. 1 and Table 1) region was amplified using 12.5 µL
 170 of NEBNext® High-fidelity 2x PCR Master Mix (New England Biolabs, Frankfurt am Main,
 171 Germany), 0.25 or 0.4 µM (Table 2) of each primer tagged with the Nanopore adapters (Fw:
 172 TTTCTGTTGGTGCTGATATTGC, Rv: ACTTGCCTGTCGCTCTATCTTC), 2-6 ng/µL of
 173 DNA (see Table 2), and nuclease-free water for a reaction of 25 µL. The thermal profile was
 174 performed with an initial denaturation at 95 °C for 5 min, 28 cycles of denaturation at 95 °C for
 175 30 s, annealing at 55 °C for 30 s, and extension at 72 °C for 60 or 90 s (see Table 2) and a final
 176 extension at 72 °C for 10 min. To inhibit plant amplification, separate reactions using ITS-PNA
 177 (Peptide Nucleic Acid) blockers (5'-CGAGGGCACGTCTGCCTGG-3' (Cregger et al. 2018))
 178 (PNA Bio Inc., Thousand Oaks, CA, USA) were tested for some combinations, adding 0,25
 179 µM of PNA and including an additional clamping step at 75 °C for 10 s before the annealing in

180 the regular thermal profile. The PCR products were purified using AMPure XP beads (Beckman
 181 Coulter Genomics, Danvers, MA, USA), followed by verification in 1% Agarose gel and
 182 quantification using the Quantifluor dsDNA system.

183



184

185 **Figure 1. Scheme of the ITS operon with the primers evaluated in this study.** The small subunit (SSU), ITS1,
 186 5.8S, ITS2, and long subunit (LSU), as well as the position of the forward primers (blue) and the reverse primers
 187 (red), are depicted.

188

189 **Table 1.** List of primers used to prepare the different combinations evaluated in this study.

Primer	Original name	Sequence	Reference
Forward	ITS1catta	ACCWGC GGARGGATCATT A	(Tedersoo and Anslan 2019)
Reverse	ITS4ngsUni	CCTSCSCTTANTDATATGC	(Tedersoo and Lindahl 2016)
Forward	ITS9munngs	TACACACCGCCCGTCG	(Tedersoo and Lindahl 2016)
Reverse	ITS4Archaeor	CCTCGCCTTATTGATATGC	(Tedersoo and Anslan 2019)
Forward	ITS1Fngs	GGTCATTTAGAGGAAGTAA	(Tedersoo et al. 2015)
Forward	ITS1ngs	TCCGTAGGTGAACCTGC	(White et al. 1990; Tedersoo et al. 2015)
Reverse	ITS4ngs	TCCTSCGCTTATTGATATGC	(White et al. 1990)
Reverse	LR5	TCCTGAGGGAAACTTCG	(Vilgalys and Hester 1990)

190

191 *Library preparation and sequencing*

192 The purified products were submitted to a second round of PCR to attach sequencing barcodes
 193 by using 1 µL of barcode from the PCR Barcoding Expansion 1-96 kit (EXP-PBC096) (Oxford
 194 Nanopore Technologies, Oxford, United Kingdom), 25 µL of LongAmp Taq 2x master mix
 195 (New England Biolabs), and 24 µL of the purified DNA (100 fmol). The products were purified
 196 using AMPure XP beads, quantified with the Quantifluor dsDNA system, and pooled in
 197 equimolar amounts to prepare a library of 1 µg. The library was submitted to an end-preparation
 198 step using the NEBNext Ultra II End Repair/dA-tailing Module (New England Biolabs),
 199 followed by an adapter ligation step using the NEBNext Quick T4 DNA Ligase (New England
 200 Biolabs) and the Ligation Sequencing Kit V14 (SQK-LSK114) (Nanopore). Finally, the
 201 sequencing runs were conducted in a MinION Mk1C using an R10.4.1 Flow Cell (FLO-
 202 MIN114) (Nanopore).

203

204 **Table 2.** Details of primer combinations evaluated in this study and the PCR conditions for amplification. The
 205 DNA volume described is based on an initial concentration between 2 and 6 ng/ μ L for a reaction of 25 μ L. The
 206 primer concentration is the final concentration of each primer in the reaction. Only the combinations in bold were
 207 used for PNA blocking and further sequencing assays.

Combination	Primer Fw	Primer Rv	Extension time (s)	DNA volume (μ L)	Primer concentration [μ M]
1	ITS1catta	ITS4ngsUni	60	5	0.4
2	ITS1catta	ITS4Archaeor	60	5	0.4
3	ITS9munngs	ITS4ngsUni	60	3	0.25
4	ITS1catta	ITS4Archaeor (50%) + ITS4ngsUni (50%)	60	5	0.4
5	ITS1catta	ITS4Archaeor (5%) + ITS4ngsUni (95%)	60	5	0.4
6	ITS1Fngs	ITS4ngsUni	60	2	0.25
7	ITS1ngs	ITS4ngsUni	60	2	0.25
8	ITS1catta	ITS4ngs	60	2	0.25
9	ITS9munngs	ITS4ngs	60	2	0.25
10	ITS1Fngs	ITS4ngs	60	2	0.25
11	ITS1ngs	ITS4ngs	60	2	0.25
12	ITS1catta	LR5	90	2	0.25
13	ITS9munngs	LR5	90	2	0.25
14	ITS1Fngs	LR5	90	2	0.25
15	ITS1ngs	LR5	90	2	0.25

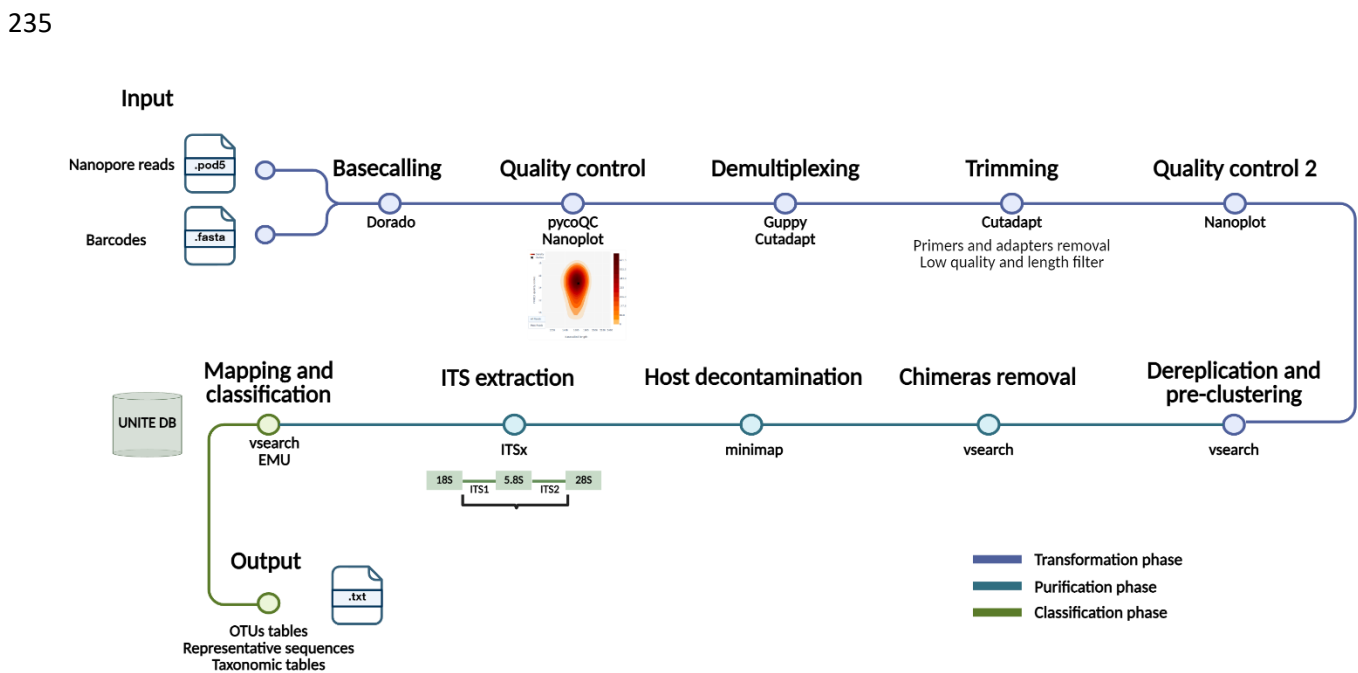
208

209 Data processing

210 *Data recovery and cleaning*

211 The raw data were recovered as pod5 files and converted to fastq (basecalling) using the duplex
 212 option of Dorado (v0.5.2) in GPU mode, obtained from the Nanopore repository
 213 (<https://community.nanoporetech.com/downloads>). The quality of the data was then inspected
 214 using pycoQC (Leger and Leonardi 2019) and NanoPlot (De Coster et al. 2018).
 215 Demultiplexing was conducted with the basecoder function of Guppy (v6.5.7) in GPU mode,
 216 followed by an additional classification round with Cutadapt (Martin 2011) for still unclassified
 217 files. Reads were filtered and trimmed using Cutadapt (Martin 2011). Only reads with a q-score
 218 higher than 10 and a length between 600 – 1000 bp for the full ITS fragments, or between 1200
 219 – 2000 bp for the fragments that included a portion of the LSU, respectively, were kept. The
 220 sequences before and after the primers were trimmed. Finally, the quality of the reads was re-
 221 inspected using NanoPlot (De Coster et al. 2018). This process is denoted as the Transformation
 222 phase (Fig. 2). Pre-clustering, decontamination, and ITS extraction

223 To facilitate classification and reduce processing power, the filtered amplicons were subjected
 224 to a dereplication and pre-clustering step using vsearch (v2.28.1) (Rognes et al. 2016). The
 225 dereplication was performed using the default parameters, while the pre-clustering was carried
 226 out using a pairwise identity of 0.97, extracting the centroids of each cluster and sorting the
 227 sequences by their decreasing abundance (cluster_size option). Subsequently, chimeras were
 228 identified *de novo* and removed using the uchime_denovo option. The rhizosphere samples
 229 were additionally submitted to a host decontamination step by mapping the reads against the
 230 genome of *Fagus sylvatica* (GCA_907173295.1) using minimap2 (v2.26-r1175) (Li 2018). The
 231 whole ITS region of all samples and primer combinations was then extracted using ITSx
 232 (v1.1.3) (Bengtsson-Palme et al. 2013). For Combination 12, only reads where the full ITS and
 233 LSU regions were found were used for further analysis. This step is denoted as the purification
 234 phase (Fig. 2).



236
 237 **Figure 2. Pipeline designed to process ITS amplicons from Nanopore sequencing.** The software and outputs
 238 of each step are depicted below.

239
 240 *Taxonomic assignment of the individual strains*

241 To confirm the identity of the strains that comprised the mock community, each strain was
 242 individually sequenced using two primer combinations (Combination 1 and 12). The reads were
 243 then submitted to the previous cleaning, pre-clustering, and ITS extraction steps, excluding host
 244 decontamination. Then, reads were sorted to obtain the consensus sequences using
 245 Amplicon_sorter (Vierstraete and Braeckman 2022) with a similarity of 98%, corrected using
 246 medaka (<https://github.com/nanoporetech/medaka>) and classified using blastn against the
 247 UNITE database (v10.0) (Abarenkov et al. 2024), for Combination 1, or the Eukaryome
 248 database (v1.9.4) (Tedersoo et al. 2024) for Combination 12, respectively.

249
 250 *Clustering, alignment, and classification of the mixed samples*

251 After the ITS extraction, the mock community and rhizosphere samples were dereplicated again
 252 and clustered with an identity of 0.97 using vsearch (Rognes et al. 2016). Then, the centroids

253 obtained after the pre-clustering step were aligned against the final centroids to calculate the
 254 abundance of the OTUs. Finally, the centroids were taxonomically assigned using the SINTAX
 255 function of vsearch against the UNITE database (v10.0) (Abarenkov et al. 2024), or the
 256 Eukaryome database (v1.9.4) (Tedersoo et al. 2024) for the full ITS + LSU primer combination,
 257 respectively. Alternatively, the classification and abundance quantification were performed
 258 using Emu (v3.4.5) (Curry et al. 2022) after the ITS extraction to compare the performance of
 259 both methods. This step was denoted as the Classification phase (Fig. 2). The whole pipeline
 260 was implemented in Snakemake as a workflow and is accessible via Github
 261 (https://github.com/Claudia-Barrera/Long_ITS_metabarcoding).

262

263 **Data analysis**

264 *Taxonomic classification of the individual strains and the mock community*

265 To assess the accuracy of the alignment during the classification of the individual strains when
 266 a short or a long ITS fragment was used against the UNITE or the Eukaryome databases, the
 267 taxonomic assignment was scored from 1 to 10 as a result of the product between the percentage
 268 of identity and the percentage of reads in the primary consensus sequence, 10 being a perfect
 269 score. Likewise, to evaluate the accuracy of classification of the mock community, the metrics
 270 true positive (TP), false positive (FP), false negative (FN), and true negatives were calculated
 271 and implemented in the Matthew's correlation coefficient (MCC) as described in Hleap et al.
 272 (2021). Briefly, the TP were defined as the number of reads assigned to a taxon known to be
 273 present in the mock community. The FP were sequences assigned to a taxon that was not present
 274 in the community, while the FN were taxa known to be present in the community, but with no
 275 sequences assigned. Finally, to calculate the TN, 10% of the sequences after the chimera
 276 removal step were sampled with Seqkit (v2.7.0) (Shen et al. 2016) and shuffled using the esl-
 277 shuffle function of HMMER3 (v3.4) (Eddy 2011) to create a set of randomized sequences that
 278 should not be assigned to any taxa. Thus, the TN corresponded to the total of randomized
 279 sequences minus the sequences assigned to any taxonomic rank. Then, the Matthews correlation
 280 coefficient (MCC) was calculated as follows (Chicco and Jurman 2023):

$$281 \quad MCC = \frac{TP \cdot TN - FP \cdot FN}{\sqrt{(TP + FP) \cdot (TP + FN) \cdot (TN + FP) \cdot (TN + FN)}}$$

282 Imbalanced data has less influence on this coefficient and it has been shown to yield more
 283 reliable results, as it considers all four categories (TP, FP, FN, TN) proportionally. The MCC
 284 has values between -1 and +1, where +1 corresponds to a perfect classification, 0 means that
 285 the prediction is not better than chance, and -1 indicates that the classification is entirely the
 286 opposite of what is observed (Chicco and Jurman 2020).

287

288 *Validation of the different approaches using mixed samples*

289 To evaluate the performance and formation of artifacts using different primers with and without
 290 the inclusion of PNAs, the percentage of retained reads after each step in the purification phase
 291 (chimera removal, host decontamination, and ITS extraction) was calculated. Moreover, the
 292 taxonomic resolution of the different primer combinations, the inclusion of PNAs, and the
 293 procedures during the purification phase were assessed by determining the fraction of reads
 294 assigned to each taxonomic rank and the composition of the samples using the phyloseq
 295 package in R (McMurdie and Holmes 2013). Finally, the richness of the major soil-related
 296 lineages (Ascomycota and Basidiomycota) was compared across the different approaches by

297 normalizing the observed richness against the square root of the sequencing depth (Tedersoo
 298 and Anslan 2019). Then, the difference between primers was calculated as the fold change from
 299 the average value across all primers. Due to the data not having a normal distribution, as
 300 indicated by the Shapiro-Wilk and Bartlett tests, a Kruskal-Wallis test with an α of 0.05 was
 301 performed to compare the statistical differences among the groups. Only groups where
 302 significant differences were found were tested for multiple comparisons using a post hoc test
 303 with Fisher's least significant difference criterion. All analyses and plots were produced using
 304 R version 4.3.1.

305

306 **Results**

307 **Identification of the individual strains**

308 To identify potential pitfalls during taxonomic classification and to evaluate the taxonomic
 309 resolution when using both short and long ITS fragments, all strains of the mock community
 310 were sequenced individually using primer combinations 1 and 12. The average length of the
 311 reads in Combination 1 was 545.1 bp, while in Combination 12 it was 1628.2 bp. Although a
 312 single consensus sequence was expected for each pure strain, an average of 1.25 consensus
 313 sequences was detected in Combination 1, and 1.82 in Combination 12, respectively (Table S2).
 314 Nevertheless, in most cases the primary consensus sequences contained more than 90% of the
 315 reads (Table S2); therefore, only the classification of these sequences was considered for
 316 identification, as given in Table 3.

317

318 The implemented score system, which combined the percentage of identity during alignment
 319 and the percentage of reads in the main consensus sequence, allowed for the identification of
 320 the effects of fragment length and database. Both options performed consistently good,
 321 averaging 9.84 out of 10 for Combination 1 and 9.67 for Combination 12 (Table 3).
 322 Interestingly, when the ITS region from Combination 12 is extracted and used without the LSU,
 323 the results of the taxonomic classification using the UNITE database are identical to those of
 324 Combination 1, with no changes in the average score.

325

326 To assess the suitability of the databases for the length of the amplified fragment, sequences
 327 from Combination 1 were classified against the UNITE database, whereas sequences from
 328 Combination 12 were classified using the Eukaryome database, which includes the complete
 329 rRNA operon. Using Combination 1 and UNITE, all strains were correctly assigned to the genus
 330 level (Table 3). In contrast, when using Combination 12 with Eukaryome, two strains
 331 (*Pleurotus ostreatus* and *Rhodotorula mucilaginosa*) were assigned to the incorrect genus,
 332 resulting in a high number of mismatches (Table S2). However, in both cases, the output
 333 belonged to the right family. Furthermore, the strain *Rhizopus microspores* could only be
 334 classified to the family rank, with 3 consensus sequences found.

335

336 At the species level, the identified taxa differed in some strains. Particularly, species identified
 337 as *Bonderzevia montana*, *Fusarium oxysporum*, *Saccharomyces eubayanus*, *Aspergillus*
 338 *oryzae*, and both species of *Cladosporium* did not match the original identification using any
 339 of the fragments (Table 3). Nevertheless, all of them obtained a high score, indicating a high
 340 percentage of identity concentrated in a unique consensus sequence. Exceptions were
 341 *Saccharomyces eubayanus* and *Aspergillus oryzae* in Combination 12 with 2 consensus

342 sequences (Tables 3 and S2). In comparison, species such as *Craterellus sp.*, *Laetiporus*
 343 *sulphureus*, and *Chlorociboria aeruginascens* scored the lowest values using Combination 12
 344 and the Eukaryome database, and accumulated a high number of additional consensus
 345 sequences (Table 3 and S2). However, the strain *Chlorociboria aeruginascens* could not be
 346 identified at the species level with any fragments or databases (Table 3).

347

348 **Table 3.** Taxonomic identification of the single strains using a short ITS fragment (Combination 1) and a long
 349 fragment (Combination 12) aligned with BLAST against the UNITE and the Eukaryome database, respectively.
 350 The classification score is depicted for each result and was calculated as the product of the percentage of identity
 351 and the percentage of reads in the main cluster on a scale from 1 to 10, where 10 is the maximum score. The
 352 highest scores are shown in green and the lowest in red.

Original ID	Comb. 1 + UNITE DB	Comb. 12 + Eukaryote DB
<i>Agaricus bisporus</i>	<i>Agaricus bisporus</i>	<i>Agaricus bisporus</i>
<i>Alternaria alternata</i> 22-2	<i>Alternaria eichhorniae</i>	<i>Alternaria sp.</i>
<i>Armillaria mellea</i>	<i>Armillaria mellea</i>	<i>Armillaria mellea</i>
<i>Aspergillus niger</i>	<i>Aspergillus piperis</i>	<i>Aspergillus niger</i>
<i>Aspergillus oryzae</i> DSM 1862	<i>Aspergillus flavus</i>	<i>Aspergillus sojae</i>
<i>Bonderzevia montana</i>	<i>Bondarzewia tibetica</i>	<i>Bondarzewia mesenterica</i>
<i>Botrytis cinerea</i>	<i>Botrytis caroliniana</i>	<i>Botrytis cinerea</i>
<i>Chlorociboria aeruginascens</i>	<i>Chlorociboria sp.</i>	<i>Chlorociboria sp.</i>
<i>Cladosporium cladosporioides</i>	<i>Cladosporium succulentum</i>	<i>Cladosporium xylophilum</i>
<i>Cladosporium montecillanum</i>	<i>Cladosporium succulentum</i>	<i>Cladosporium xylophilum</i>
<i>Craterellus sp.</i>	<i>Craterellus tubaeformis</i>	<i>Craterellus cornucopioides</i>
<i>Fomes fomentarius</i>	<i>Fomes fomentarius</i>	<i>Fomes fomentarius</i>
<i>Fusarium oxysporum</i>	<i>Fusarium mori</i>	<i>Fusarium bostrycoides</i>
<i>Ganoderma applanatum</i>	<i>Ganoderma applanatum</i>	<i>Ganoderma oregonense</i>
<i>Hericium coralloides</i>	<i>Hericium coralloides</i>	<i>Hericium coralloides</i>
<i>Hypsizygus tessulatus</i>	<i>Hypsizygus marmoreus</i>	<i>Hypsizygus marmoreus</i>
<i>Laetiporus sulphureus</i> (mycelium)	<i>Laetiporus ailaoshanensis</i>	<i>Laetiporus sulphureus</i>
<i>Laetiporus sulphureus</i> (fruiting body)	<i>Laetiporus sulphureus</i>	<i>Laetiporus sulphureus</i>
<i>Lentinula edodes</i>	<i>Lentinula novae-zelandiae</i>	<i>Lentinula edodes</i>
<i>Penicillium roqueforti</i>	<i>Penicillium roqueforti</i>	<i>Penicillium roqueforti</i>
<i>Pleurotus ostreatus</i>	<i>Pleurotus ostreatus</i>	<i>Hohenbuehelia auriscalpium</i>
<i>Rhizopus microsporus</i>	<i>Rhizopus microsporus</i>	Rhizopodaceae
<i>Rhodotorula mucilaginosa</i>	<i>Rhodotorula mucilaginosa</i>	<i>Rhodospordiobolus mucilaginosa</i>
<i>Saccharomyces eubayanus</i> CBS 12357	<i>Saccharomyces bayanus</i>	<i>Saccharomyces uvarum</i>
<i>Sordaria macrospora</i> R19027	<i>Sordaria fimicola</i>	<i>Sordaria macrospora</i>
<i>Trametes versicolor</i> BAM116 (CTB863)	<i>Trametes versicolor</i>	<i>Trametes versicolor</i>

353

354 **Accuracy of the primers and methods**

355 Of the 15 primer combinations initially tested (Table 2), six were finally maintained based on
 356 their ability to produce high-quality amplicons. The performance of these six primer
 357 combinations was further evaluated based on the number of reads obtained, their q-score, and
 358 the number of reads retained after chimera removal and ITS extraction steps (Table 4). Reads
 359 from Combinations 1, 7, 8, and 10 with primers flanking the ITS region had an average length
 360 between 693.4 and 772.9 bp. Combinations 3 and 12 that include part of the SSU or the LSU
 361 resulted in an average length of 901.7 and 1631.4 bp, respectively. Although no differences
 362 were found regarding the q-score, Combination 12 yielded an elevated number of reads,
 363 accompanied by a substantial accumulation of chimeras and aberrant reads, resulting in a very
 364 low ITS extraction rate (Table 4). On the contrary, primer combinations for shorter fragments
 365 were less susceptible to accumulating chimeras. However, the extraction rates of ITS varied
 366 strongly between combinations, with Combinations 1 and 8 retaining only 25.5% and 29.3% of
 367 reads after the ITS extraction, whereas Combinations 3, 7, and 10 showed a retention rate higher
 368 than 70%.

369

370 **Table 4.** Properties of the different primer combinations used for assessing the mock community. Read lengths,
 371 numbers of initial reads, and q-scores were calculated after the second quality control during the Transformation
 372 phase (Fig. 2).

Combination	Read length (bp)	No. Reads	q-score	Reads retained (%)	
				Chimera removal	ITS extraction
1	693.4	17424	14.8	87.3%	25.5%
3	901.7	26736	15.2	85.1%	80.0%
7	772.9	16526	15.5	91.6%	70.3%
8	721.3	12344	15.4	88.6%	29.3%
10	750.2	14966	15.3	84.5%	83.3%
12	1631.4	88504	15.6	72.8%	5.9%

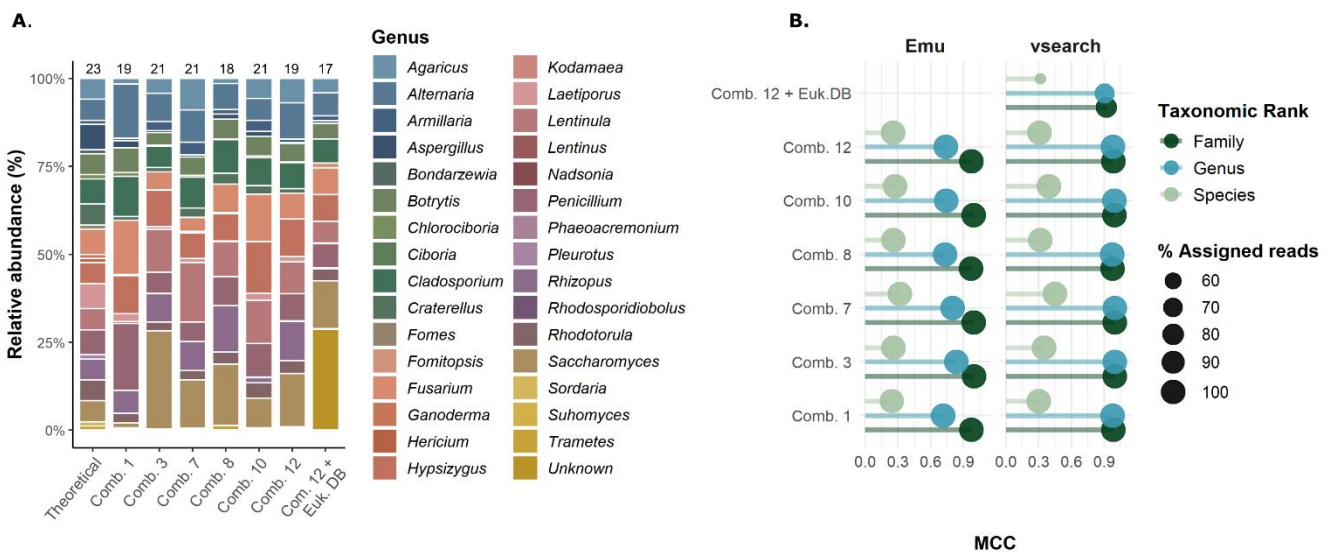
373

374 Although no primer combination was able to recover all 23 genera present in the mock
 375 community, Combinations 3, 7, and 10 recovered the highest number of genera with both
 376 classifiers (21 for vsearch, 16 for EMU) (Figs 3A, S1). The genera in the lowest concentration
 377 in the original mock community were difficult to recover with the different Combinations. For
 378 instance, genera such as *Bondarzewia* could not be identified in any combination or classifier
 379 except in low abundance in Combination 12 against the Eukaryome database. The Emu
 380 classifier failed to recognize reads of the genera *Botrytis* (assigned as *Botryotina*), *Fomes*,
 381 *Ganoderma*, *Hericium*, *Pleurotus*, or *Sordaria* with any of the primer combinations (Fig. S1),
 382 which were also present in low concentration in the original community. On the contrary, the
 383 vsearch classifier was more sensitive to low-abundance genera, despite groups such as *Fomes*,
 384 *Hericium*, and *Pleurotus* being absent in half of the primer combinations.

385

386 Around 99% of the reads could be assigned up to the species level with all the primer
 387 combinations and classifiers, except for those classified against the Eukaryome database, where
 388 only 51% could be identified (Fig. 3B). Matthew's correlation coefficient (MCC) was
 389 calculated to evaluate the accuracy of the assignment. Both classifiers demonstrated high
 390 classification accuracy at the family rank, with an MCC above 0.97; however, only vsearch
 391 maintained this performance at the genus level (Fig. 3B). By comparison, the accuracy at the

392 species rank was between 0.3 and 0.45 for vsearch and 0.19 and 0.30 for EMU, which agrees
 393 with the results obtained during the classification of the individual strains. Although the
 394 classification at the species level was not perfect, the positive values indicate that it was not a
 395 product of chance.



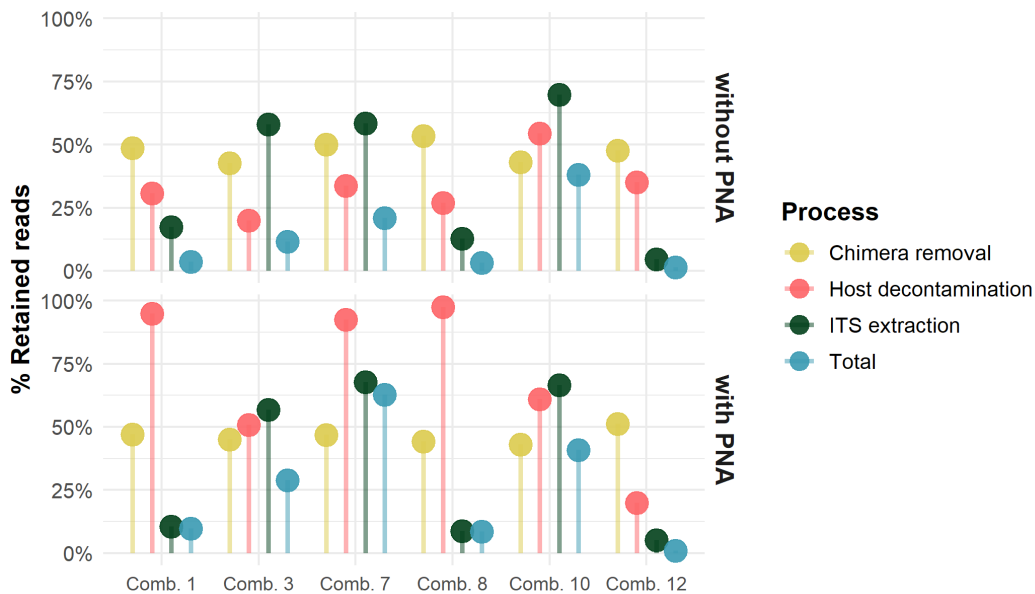
397 **Figure 3. Community composition and accuracy, as determined by the different primer combinations and**
 398 **classifiers used to assess the mock community. A.** Taxonomic composition at the genus level of the mock
 399 community using vsearch for classification. The theoretical composition is based on the amount of DNA added
 400 during mock community preparation. The numbers over the stacked bars denote the unique genera found in each
 401 combination and known to be present in the theoretical community. The results from Combination 12 were
 402 classified against the UNITE and Eukaryome databases (Euk. DB). **B.** Results of the Matthews correlation
 403 coefficient (MCC) were used as a parameter to evaluate the classification accuracy at the family, genus, and species
 404 levels.

405
 406 **Resolution power and variability using rhizosphere samples**

407 Applying and testing the pipeline with rhizosphere samples showed similar percentages of reads
 408 recovered after chimera removal with all primer combinations (Fig. 4, Table S3). However, the
 409 differences were more notable after the host decontamination step, where the retention without
 410 PNAs was between 20% for Combination 3 and 54.3% for Combination 10. Including PNA,
 411 retention was substantially improved to almost 95% for Combinations 1, 7, and 8, while in
 412 Combination 12, it seemed to promote plant ITS amplification. The ITS extraction yield was
 413 above 57% in Combinations 3, 7, and 10 with or without PNAs, while it was low for
 414 Combinations 1 and 8, similar to what was observed for the processing of the mock community.
 415 This demonstrated that including PNA successfully inhibited the amplification of the plant ITS,
 416 improving the retention of total reads for most combinations. Moreover, Combinations 7 and
 417 10 exhibited the highest retention of total reads, in the case of Combination 10, even when no
 418 PNA was included.

419
 420 The fraction of reads assigned to each taxonomic rank was determined to evaluate the resolution
 421 power of the taxonomic classification with the different methods. Almost 100% of the retained
 422 reads in all combinations could be assigned using vsearch, whereas with Emu, up to 25% of the
 423 reads in Combination 12 remained unclassified (Fig. 5A). Although not all the reads could be
 424 assigned using Emu, around 50% of classified reads could be assigned to the species level under
 425 any of the modifications (adding PNAs during PCR or purifying the reads). Additionally, a drop

426 in the number of reads assigned to the species level was observed when PNA was included,
 427 data were purified, or both, independently of the classifier. This reduction is the result of the
 428 depletion of plant ITS sequences in the samples (Figs 5B, S2), considering that most of the
 429 plant reads were classified to the species level.

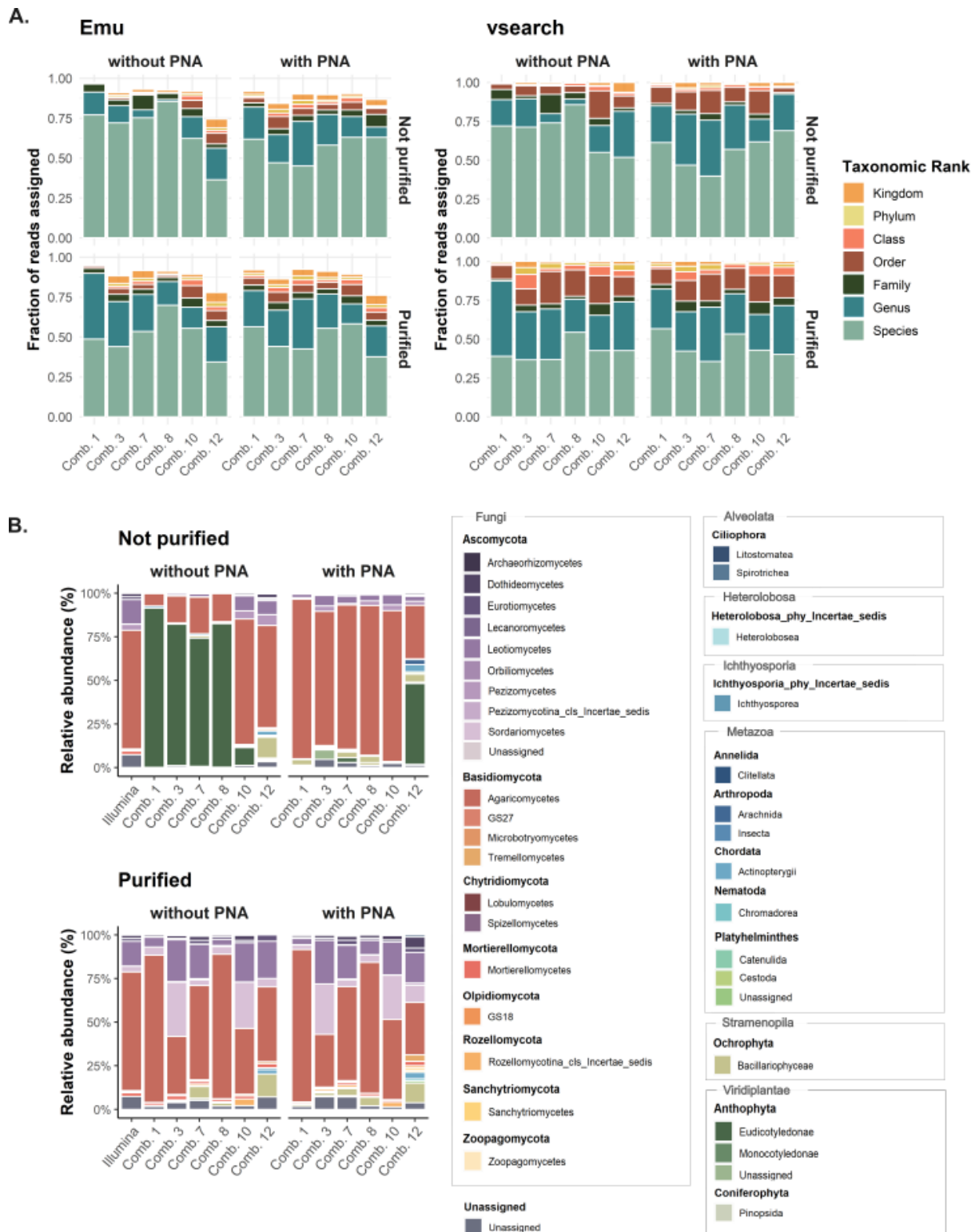


430

431 **Figure 4. Percentage of reads retained using different experimental and purification steps.** The proportion
 432 of retained reads, with and without PNA, was compared across various purification steps during amplification.
 433 The previous step served as a reference to calculate the number of reads removed, except in the Total values, where
 434 the final number of reads was compared with the number before purification. Detailed values are presented in
 435 Table S3.

436

437 The fungal composition without PNA or sequence purification was masked by the dominating
 438 presence of plant ITS, particularly of the class Eudicotyledonae family Fagaceae, in almost all
 439 combinations except for 10 and 12 (Figs 5B, S2). This strongly points to interference caused
 440 by the host and the need for a strategy to inhibit the contamination. Our results showed that the
 441 plant contamination ranged between 10% for Combination 10, to 90% for Combination 1. The
 442 contamination was successfully removed using the sequence purification step, with the
 443 disadvantage of eliminating between 65% to 80% of the reads (Fig. 4). These values were
 444 improved by adding PNA, except for Combination 12, where the relative abundance of
 445 Eudicotyledonae increased by c. 50% when PNA was included, but reads purification was
 446 omitted. The use of PNA without the purification step increased the abundance of fungal reads
 447 by 12-fold for Combination 1, 4.1-fold for Combination 3, 2.9-fold for Combination 7, 4.7-fold
 448 for Combination 8, and 0.1-fold for Combination 10. In contrast, the abundance for
 449 Combination 12 decreased by 0.5-fold (Fig. 5B). The inclusion of a sequence purification step
 450 revealed more members of the Ascomycota phylum across all primer combinations, particularly
 451 when vsearch was used as a classifier. Remarkably, the class Archaeorhizomycetes was
 452 recovered when data were purified in Combinations 3, 8, and 10, and in Combination 1 when
 453 PNA was also included. Contrastingly, the class Pezizomycetes, despite not being identified by
 454 Illumina, was recovered in Combinations 3, 7, and 8 after sequence purification, and in
 455 Combinations 1, 8, and 12 after PNA inclusion and sequence purification. Moreover, classes
 456 such as Lecanoromycetes, Orbiliomycetes, and Sordariomycetes increased in abundance in
 457 almost all combinations with the purification step, independently of adding PNA, except in
 458 Combination 12.

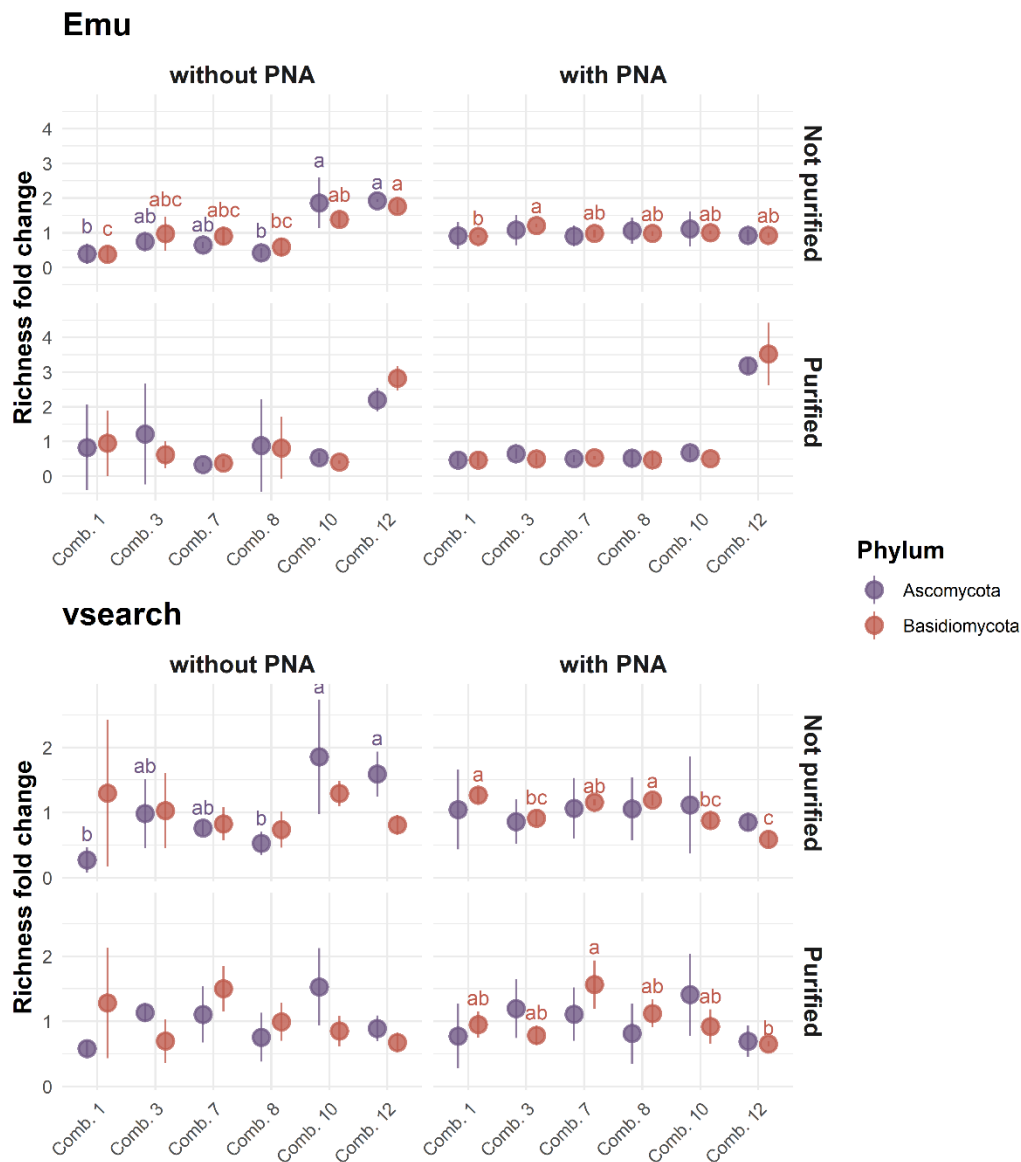


459

460 **Figure 5. Taxonomic assignment under different experimental and bioinformatic strategies.** **A.** Proportion
 461 of reads assigned to each taxonomic rank under various primer combinations, classifiers (Emu and vsearch), PNA
 462 inclusion, and reads purification (host decontamination and ITS extraction). **B.** Community composition at the
 463 class level of one of the rhizosphere samples using vsearch as a classifier. The legend depicts Kingdoms in gray
 464 boxes, phyla with bold subtitles, and the internal classes. Results obtained using Illumina (MiSeq v3) with the
 465 amplification of the ITS2 fragment are also included for comparison. The evaluation using other samples showed
 466 similar patterns.

467

468 The sample composition results obtained with Illumina were more similar to the Emu results
 469 compared to those obtained with vsearch, particularly when PNA and purification were
 470 included (Fig. S2). However, using Emu as a classifier also resulted in more unclassified reads,
 471 which can be resolved by utilizing vsearch instead. Moreover, vsearch not only reduced the
 472 number of unclassified sequences but also highlighted the differences between some primer
 473 combinations, enabling the identification of taxa from various protist groups, such as Alveolata,
 474 particularly in Combination 7, and Metazoa (Fig. 5B) in almost all primer combinations.
 475



476

477 **Figure 6. Differences in the richness of the Phyla Ascomycota and Basidiomycota using different**
 478 **experimental and bioinformatic approaches.** The x-axis shows the primer combinations, and the y-axis shows
 479 the fold change in richness relative to the average across all primers. The error bars correspond to the standard
 480 deviation. A Kruskal-Wallis test with an α level of 0.05 was conducted to assess differences between groups. Only
 481 the groups with statistically significant differences were tested for multiple comparisons using a post hoc test with
 482 Fisher's least significant difference criterion. The letters above the bars indicate the significant differences in the
 483 multiple-comparison test.

485 The richness of taxa assigned to the Ascomycota and Basidiomycota phyla was more variable
 486 using vsearch compared with Emu. Adding PNA and classifying with Emu homogenized the
 487 richness between different primer combinations, except for Combination 12, where the richness
 488 values increased 2- to 3-fold after purification. When using vsearch instead of Emu, the pattern
 489 of richness of the Basidiomycota and Ascomycota phyla remained similar, regardless of
 490 whether PNA was added. However, vsearch results underwent some changes when the data
 491 were subjected to purification. For example, the richness of Basidiomycota in Combinations 3
 492 and 7 experienced a drop or an increase in fold change, respectively, when the data were
 493 purified. Additionally, the statistical difference found between primer combinations in
 494 Ascomycota when vsearch is used, but not PNA or data purification, disappeared with the
 495 inclusion of these modifications. Nevertheless, the richness of the Basidiomycota remained
 496 significantly different between the primer combinations when PNA was added independently
 497 of whether the data were purified or not (Fig. 6).

498

499 Discussion

500 Identification of individual strains

501 Identifying the individual cultured strains using the same forward primer, ITS1catta, and two
 502 different reverse primers (ITS4ngsUni and LR5) allowed for the discovery of potential issues
 503 during classification when either short or long fragments were used. As a first observation, an
 504 increment of 31.4% in the number of consensus sequences was found for the long fragment
 505 compared with the short fragment. This difference suggests that the length of the fragment may
 506 influence the accumulation of errors during amplification and that these errors cannot be
 507 corrected during data processing, resulting in clusters of additional consensus sequences.
 508 Indeed, the rate of mismatch accumulation during the alignment was 25.3% higher in the long
 509 fragment than in the short one (Table S2). These results align with those presented by Ohta et
 510 al. (2023), which showed that the number of mismatches increased with the length of the
 511 fragment, although this did not affect their final classification result. Consequently, despite a
 512 slight increase in the number of consensus sequences when the long fragment was used, our
 513 results showed that the primary sequence accumulated more than 95% of the reads, and the
 514 classification using this sequence had a percentage of identity of around 99% during the
 515 alignment in most cases (Table S2).

516

517 Although both fragments and databases accurately identified most of the strains at the genus
 518 level, except for two cases using the long fragment with the Eukaryome database, the results
 519 were less consistent at the species level. There, 12 strains out of 26 had an exact match using
 520 either the short or the long fragment, while species within the genera *Fusarium*, *Saccharomyces*,
 521 *Cladosporium*, or *Aspergillus* could not be matched using both fragments and databases. This
 522 likely relates to the complex evolutionary process of some fungal groups, which is generally
 523 challenging to elucidate with the ITS region alone (Lücking et al. 2020). In this regard, many
 524 authors agree that the correct identification up to species level is favored by evaluating
 525 additional markers, such as calmodulin, β -tubulin, RPB2, or the translation elongation factor 1
 526 (tef1) (Watanabe et al. 2011; Samson et al. 2014; Lee et al. 2023). Moreover, cases as *Fusarium*
 527 *oxysporum*, *Cladosporium cladosporioides*, or *Cladosporium herbarum* are considered species
 528 complexes with more than one species assigned, whose identification can only be solved with
 529 a multilocus DNA sequence approach (Schubert et al. 2007; Bensch et al. 2010; Pastrana et al.
 530 2017). Therefore, imprecise classifications of these species were expected, although their
 531 correct assignment within the species complex was always achieved.

532

533 The results of the classification scores and the taxonomic identification indicated that a
 534 fragment comprised of the full ITS and part of the LSU (Combination 12) does not outperform
 535 the results obtained using only the full ITS (Combination 1) when individual strains are
 536 identified. Herein, Heeger et al. (2018) found that classification using the ITS was more
 537 frequently assigned to the species level than the SSU or the LSU alone. Likewise, Ohta et al.
 538 (2023) stated that no changes in taxonomic resolution were found by including the latter two-
 539 thirds of the LSU fragment, presumably because there is insufficient variability in this region
 540 (Schoch et al. 2012). Nevertheless, the potential of identification using the ITS in combination
 541 with the SSU and the LSU could be hidden by the limited number of sequences in the database
 542 dedicated to collecting the full operon.

543

544 **Accuracy of the primers and the classification system**

545 Using a mock community, we evidenced potential biases across different primer combinations
 546 and steps during the classification process. The proportions of chimeras ranged from 8.4% to
 547 15.5% in all primer combinations, except for Combination 12, which had a proportion of 27.2%.
 548 These values agree with those reported by Heeger et al. (2018) for an amplification process
 549 with more than 25 cycles. According to the authors, the formation of chimeras could be reduced
 550 if the PCR cycles were kept between 13 and 18. However, this strategy was not explored in our
 551 study. Interestingly, the ITS extraction yield seemed to be correlated with the forward primer
 552 since combinations 1, 8, and 12, which used ITS1catta as a forward primer, showed the lowest
 553 proportion of reads after using ITSx and identified the lowest number of genera in the mock
 554 community with any of the classifiers. This primer was designed to bypass the SSU 3' intron,
 555 enabling the recovery of a large fungal diversity without plant interference (Tederloo and
 556 Anslan 2019). Nevertheless, Winand et al. (2025) reported that ITSx did not recognize the first
 557 bases of the SSU included in the 5'-end of the ITS1catta primer. Similar results were
 558 documented by Tederloo et al. (2018) when primers flanking the ITS2 region were compared
 559 against the ITS1 or the full ITS. In that case, the authors hypothesized that the HMM algorithm
 560 of ITSx poorly recognizes the 3'-extreme of the 5.8S compared with the 5'-extreme and
 561 suggested modifying the E-value threshold in the ITSx parameters from $1e^{-5}$ to $1e^{-2}$. However,
 562 no improvements were found in our study after these adjustments.

563

564 Given the results obtained with individual strains, the classification with a combined mock
 565 community was expected to have comparable discrepancies at the species level. While the
 566 Matthews correlation coefficient (MCC) at the family and genus level showed accuracy results
 567 above 0.97, particularly using vsearch as a classifier, the results at the species level, although
 568 positive, were considerably lower. These misclassified species (in high concentration in the
 569 prepared mock community) either belong to the above discussed species complexes (*Fusarium*,
 570 *Cladosporium*, or *Aspergillus*), were assigned to outdated nomenclature, as with *Botrytis*, which
 571 was mistaken as *Botryotina* by Emu, or have an otherwise close phylogenetic relationship that
 572 is difficult to entangle, as *Hypsizygus tessulatus* (Angelini et al. 2023), identified with any
 573 primer combination or classifiers as *H. marmoreus*. Nevertheless, these results showed the
 574 potential of Combinations 3, 7, and 10 to recover more community members, and the vsearch
 575 classifier as a more sensitive and accurate method compared to Emu.

576

577 **A tailored workflow and pipeline for rhizosphere samples**

578 After establishing our primer setup and data processing with the mock community, we tested
579 them on environmental root-rhizosphere samples. One of the first complications identified was
580 the persistent co-amplification of plant ITS with most primer combinations. Despite the forward
581 primer ITS1catta being described to exclude the amplification of vascular plants under stringent
582 conditions (Tedersoo and Anslan 2019), no optimal conditions were found to guarantee the
583 amplification of only fungal ITS. Therefore, additional steps were adopted during amplification
584 and data processing, primarily involving the addition of PNA blockers to inhibit the
585 amplification of plant ITS, along with a host decontamination step that involved mapping the
586 reads against the reference genome of *Fagus sylvatica*.

587

588 Few studies have been published using ITS-PNA for short-read amplification (Cregger et al.
589 2018; Whitaker 2025), but to our knowledge, this is the first report of its implementation using
590 long-read amplification and third-generation sequencing. ITS-PNA successfully excluded plant
591 co-amplification for Combinations 1, 3, 7, and 8, improving the percentage of reads retained
592 after host decontamination. In contrast, Combination 10 exhibited a consistent amplification of
593 the fungal ITS, with minor interferences from the plant, either with or without the inclusion of
594 PNA, making this combination a promising solution in cases where PNA cannot be afforded.
595 Contrasting results were found using the long fragment amplified by Combination 12. In that
596 case, the host co-amplification was enhanced by including PNA, hinting at a disruption between
597 the blocker and the non-plant fragment. Although the host reads were successfully removed
598 after the purification steps, this combination showed the lowest yield of ITS sequence
599 extraction, suggesting that it could be affected by both the length of the fragment and the
600 recognition of the 5'-end in the forward primer ITS1catta.

601

602 In addition to the suitability of the primer selection and data processing approaches, we also
603 evaluated the level of taxonomic resolution achieved with the different approaches and their
604 ability to describe the community composition. Here, Emu and SINTAX (the primary classifier
605 of vsearch) showed some disparities in taxonomic resolution and community composition. For
606 Emu, the fraction of reads assigned to species was higher than for vsearch using the same
607 database, although a few reads could not be assigned to any taxonomic rank. Additionally, a
608 higher richness was observed with vsearch compared with Emu, both in the mock community
609 and in the rhizosphere samples. These differences are likely explained by the two-step algorithm
610 developed by Emu, where the reads are initially aligned to the reference database and then
611 corrected based on the errors expected for Nanopore sequencing (Curry et al. 2022). Despite
612 this algorithm having proven successful for 16S classification, our results for fungal
613 classification of both a mock community and environmental samples indicate that Emu could
614 underestimate the number of recovered taxa, likely because some are identified as errors and
615 removed during the second step. Therefore, the classification based on *k-mer* similarity used by
616 vsearch (Rognes et al. 2016) may serve as a better solution for recovering different taxa.
617 Moreover, the cutoff option enabled by vsearch can improve the accuracy of classification, as
618 described by Hu et al. (2022), which partially explains the higher values of MCC obtained with
619 vsearch compared to Emu.

620

621 Finally, considering the high abundance of Ascomycota and Basidiomycota phyla in
622 environmental samples, we compared the richness of both groups under different primer
623 combinations and data cleaning approaches. Our results suggest an improvement in the
624 recovery of Ascomycota after sequence purification, most likely because of the ITS extraction.

625 This recovery was also favored by primers located upstream of the 3' end of the SSU, as is the
 626 case for primer combinations 3 and 10. Although the implementation of degenerate primers,
 627 such as ITS1catta or ITS4ngsUni, has been suggested to differentiate between most fungal
 628 groups (Tedersoo et al. 2018; Tedersoo and Anslan 2019), we found a detriment in the richness
 629 and abundance of the Ascomycota phylum using this combination of primers (Combination 1).
 630 However, combining one degenerate and one universal primer (as in Combinations 3 or 7), or
 631 two universal primers (as in Combination 10), could be beneficial in improving the recovery of
 632 most fungal groups.

633

634 **Conclusion**

635 We evaluated the potential of different primer combinations and data processing approaches to
 636 implement Nanopore sequencing for profiling fungal communities in soil and the rhizosphere.
 637 Increasing the fragment length from whole ITS to whole ITS + LSU did not improve the
 638 taxonomic resolution to which individual strains were identified, despite using a database with
 639 the complete rRNA. Using a mock community, we showed that the combination of non-
 640 degenerate forward primers with universal or degenerative reverse primers retained more than
 641 70% of the reads after ITS extraction. Moreover, our pipeline identified 91% of the strains in
 642 the mock community with an MCC higher than 0.97 at the genus level, using vsearch as a
 643 classifier. Our approaches to amplify complex samples by adding PNA completely inhibited
 644 co-amplification of plant ITS in most of the primer combinations evaluated. The comparison
 645 between adding PNA and/or implementing sequence purification steps (chimera removal, host
 646 decontamination, and ITS extraction) revealed differences among the primer combinations and
 647 the benefits of using the different amplification and data processing strategies. Thus, to retain
 648 the largest number of reads and recover most of the fungal communities, we suggest working
 649 with a combination of primers such as Combination 3, 7, or 10, adding PNA when samples are
 650 susceptible to co-amplify plant-ITS, implementing host decontamination (if needed) and ITS
 651 extraction steps to recover most of the fungal diversity, and using vsearch to improve the
 652 resolution of identification. Furthermore, our modular pipeline is available and can be
 653 implemented by adding or removing steps according to the user's needs.

654

655 **Acknowledgement**

656 We thank Prof. Dr. Philip Benz, Sabrina Klasen, and Petra Arnold from the Group of Fungal
 657 Biotechnology in Wood Science (TUM School of Life Sciences, Technical University of
 658 Munich, Freising, Germany) for providing most of the isolates used for the mock community.

659

660 **Data accessibility**

661 The pipeline and all associated codes are available on GitHub: https://github.com/Claudia-Barrera/Long_ITS_metabarcoding
 662

663 Demultiplexed sequences have been deposited at the European Nucleotide Archive (ENA)
 664 under Project Accession number PRJEB106608.

665

666 **Author contribution**

667 Conceptualization: C.B., K.P., and F.W. Methodology, software, and formal analysis: C.B.
 668 Writing - Original draft: C.B. Writing - Review and Editing; all authors.

669

670 Funding information

671 Bavarian State Ministry of Food, Agriculture, Forestry, and Tourism (klifW022).

672 German Federal Ministry of Agriculture, Food, and Regional Identity (2224NR084X;

673 Fachagentur Nachwachsende Rohstoffe, Förderprogramm Nachhaltige Erneuerbare

674 Ressourcen).

675

676 Competing interests

677 The author has declared that no competing interests exist.

678

679

680 References

- 681 Abarenkov K, Henrik Nilsson R, I-Henrik Larsson K, Taylor AFS, May TW, Guldberg Frøslev T, Pawłowska J, Lindahl B,
 682 Põldmaa K, Truong C, Vu D, Hosoya T, Niskanen T, Piirmann T, Ivanov F, Zirk A, Peterson M, Cheeke TE, Ishigami Y,
 683 Tobias Jansson A, Stjernegård Jeppesen T, Kristiansson E, Mikryukov V, Miller JT, Oono R, Francisco Ossandon FJ,
 684 Dupérion JP, Saar I, Schigel D, Suija A, Tedersoo L, Kõljalg U (2024) The UNITE database for molecular
 685 identification and taxonomic communication of fungi and other eukaryotes: sequences, taxa and classifications
 686 reconsidered. *Nucleic Acids Research* 52: 791–797. doi:10.1093/nar/gkad1039.
- 687 Akter S, Mahmud U, Shoumik BAA, Khan MZ (2025) Although invisible, fungi are recognized as the engines of a microbial
 688 powerhouse that drives soil ecosystem services. *Archives of Microbiology* 207: doi:10.1007/s00203-025-04285-4.
- 689 Angelini P, Flores GA, Cusumano G, Venanzoni R, Pellegrino RM, Zengin G, Simone SCD, Menghini L, Ferrante C, Angelini P,
 690 Flores GA, Cusumano G, Venanzoni R, Pellegrino RM, Zengin G, Di Simone SC, Menghini L, Ferrante C (2023)
 691 Bioactivity and Metabolomic Profile of Extracts Derived from Mycelial Solid Cultures of *Hypsizygus marmoreus*.
 692 *Microorganisms* 11: doi:10.3390/microorganisms11102552.
- 693 Bender SF, Plantenga F, Neftel A, Jocher M, Oberholzer H-R, Köhl L, Giles M, Daniell TJ, Van Der Heijden MGA (2014)
 694 Symbiotic relationships between soil fungi and plants reduce N₂O emissions from soil. *The ISME Journal* 8: 1336–
 695 1345. doi:10.1038/ismej.2013.224.
- 696 Bengtsson-Palme J, Ryberg M, Hartmann M, Branco S, Wang Z, Godhe A, De Wit P, Sánchez-García M, Ebersberger I, De
 697 Sousa F, Amend A, Jumpponen A, Unterseher M, Kristiansson E, Abarenkov K, Bertrand YJK, Sanli K, Eriksson KM,
 698 Vik U, Veldre V, Nilsson RH (2013) Improved software detection and extraction of ITS1 and ITS2 from ribosomal
 699 ITS sequences of fungi and other eukaryotes for analysis of environmental sequencing data. *Methods in Ecology
 700 and Evolution* 4: 914–919. doi:10.1111/2041-210x.12073.
- 701 Bensch K, Groenewald JZ, Dijksterhuis J, Starink-Willemsse M, Andersen B, Summerell BA, Shin HD, Dugan FM, Schroers HJ,
 702 Braun U, Crous PW (2010) Species and ecological diversity within the *Cladosporium cladosporioides* complex
 703 (Davidiellaceae, Capnodiales). *Studies in Mycology* 67: 1–94. doi:10.3114/sim.2010.67.01.
- 704 Buetas E, Jordán-López M, López-Roldán A, D'Auria G, Martínez-Priego L, De Marco G, Carda-Diéguez M, Mira A, Buetas E,
 705 Jordán-López M, López-Roldán A, D'Auria G, Martínez-Priego L, De Marco G, Carda-Diéguez M, Mira A (2024) Full-
 706 length 16S rRNA gene sequencing by PacBio improves taxonomic resolution in human microbiome samples. *BMC
 707 Genomics* 2024 25:1 25: doi:10.1186/s12864-024-10213-5.
- 708 Butler I, Turner O, Mohammed K, Akhtar M, Evans D, Lambourne J, Harris K, O'Sullivan DM, Sergaki C (2025) *Frontiers |*
 709 *Standardization of 16S rRNA gene sequencing using nanopore long read sequencing technology for clinical*
 710 *diagnosis of culture negative infections. Frontiers in Cellular and Infection Microbiology* 15:
 711 doi:10.3389/fcimb.2025.1517208.
- 712 Chicco D, Jurman G (2020) The advantages of the Matthews correlation coefficient (MCC) over F1 score and accuracy in
 713 binary classification evaluation. *BMC Genomics* 2019 21:1 21: doi:10.1186/s12864-019-6413-7.
- 714 Chicco D, Jurman G (2023) The Matthews correlation coefficient (MCC) should replace the ROC AUC as the standard metric
 715 for assessing binary classification. *BioData Mining* 2023 16:1 16: doi:10.1186/s13040-023-00322-4.
- 716 Cregger MA, Veach AM, Yang ZK, Crouch MJ, Vilgalys R, Tuskan GA, Schadt CW, Cregger MA, Veach AM, Yang ZK, Crouch MJ,
 717 Vilgalys R, Tuskan GA, Schadt CW (2018) The *Populus* holobiont: dissecting the effects of plant niches and
 718 genotype on the microbiome. *Microbiome* 2018 6:1 6: doi:10.1186/s40168-018-0413-8.
- 719 Cuber P, Choonee D, Geeves C, Salatino S, Creedy TJ, Griffin C, Sivess L, Barnes I, Price B, Misra R (2023) Comparing the
 720 accuracy and efficiency of third generation sequencing technologies, Oxford Nanopore Technologies, and Pacific
 721 Biosciences, for DNA barcode sequencing applications. *Ecological Genetics and Genomics* 28:
 722 doi:10.1016/j.egg.2023.100181.

- 723 Curry KD, Wang Q, Nute MG, Tyshaieva A, Reeves E, Soriano S, Wu Q, Graeber E, Finzer P, Mendling W, Savidge T, Villapol S,
724 Dilthey A, Treangen TJ (2022) Emu: species-level microbial community profiling of full-length 16S rRNA Oxford
725 Nanopore sequencing data. *Nature Methods* 2022 19:7 19: 845–853. doi:10.1038/s41592-022-01520-4.
- 726 De Coster W, D'Hert S, Schultz DT, Cruts M, Van Broeckhoven C (2018) NanoPack: visualizing and processing long-read
727 sequencing data. *Bioinformatics* 34: 2666–2669. doi:10.1093/bioinformatics/bty149.
- 728 De Menezes AB, Richardson AE, Thrall PH (2017) Linking fungal–bacterial co-occurrences to soil ecosystem function. *Current*
729 *Opinion in Microbiology* 37: 135–141. doi:10.1016/j.mib.2017.06.006.
- 730 Eddy SR (2011) Accelerated Profile HMM Searches. *PLOS Computational Biology* 7: doi:10.1371/journal.pcbi.1002195.
- 731 Goodwin S, McPherson JD, McCombie WR, Goodwin S, McPherson JD, McCombie WR (2016) Coming of age: ten years of
732 next-generation sequencing technologies. *Nature Reviews Genetics* 2016 17:6 17: doi:10.1038/nrg.2016.49.
- 733 Heeger F, Bourne EC, Baschien C, Yurkov A, Boyke B, Spröer C, Overmann J, Mazzoni CJ, Michael, Monaghan T (2018) Long-
734 read DNA metabarcoding of ribosomal RNA in the analysis of fungi from aquatic environments.
735 doi:10.1111/1755-0998.12937.
- 736 Heym M, Pritsch K, Schmied G, Grams TEE, Pretzsch H, Roth M, Weigl F (2023) Diversität wurzellozierter Pilze in Rein-
737 und Mischbestand von Kiefer und Buche. *LWF aktuell*, 19–21 pp.
- 738 Hleap JS, Littlefair JE, Steinke D, Hebert PDN, Cristescu ME (2021) Assessment of current taxonomic assignment strategies
739 for metabarcoding eukaryotes. *Molecular Ecology Resources* 21: doi:10.1111/1755-0998.13407.
- 740 Ito ZA, Reshi ZA (2013) The Multifunctional Role of Ectomycorrhizal Associations in Forest Ecosystem Processes. *The*
741 *Botanical Review* 79: 371–400. doi:10.1007/s12229-013-9126-7.
- 742 Lee W, Kim JS, Seo CW, Lee JW, Kim SH, Cho Y, Lim YW, Lee W, Kim JS, Seo CW, Lee JW, Kim SH, Cho Y, Lim YW, Lee W, Kim
743 JS, Seo CW, Lee JW, Kim SH, Cho Y, Lim YW (2023) Diversity of Cladosporium (Cladosporiales, Cladosporiaceae)
744 species in marine environments and report on five new species. *MycKeys* 98: doi:10.3897/mycokeys.98.101918.
- 745 Leger A, Leonardi T (2019) pycoQC, interactive quality control for Oxford Nanopore Sequencing. *Journal of Open Source*
746 *Software* 4: 1236–1236. doi:10.21105/JOSS.01236.
- 747 Li H (2018) Minimap2: pairwise alignment for nucleotide sequences. *Bioinformatics* 34: 3094–3100.
748 doi:10.1093/bioinformatics/bty191.
- 749 Lu J, Zhang X, Zhang X, Wang L, Zhao R, Liu XY, Liu X, Zhuang W, Chen L, Cai L, Wang J (2023) Nanopore sequencing of full
750 rRNA operon improves resolution in mycobiome analysis and reveals high diversity in both human gut and
751 environments. *Molecular Ecology* 32: 6330–6344. doi:10.1111/MEC.16534.
- 752 Lücking R, Aime MC, Robbertse B, Miller AN, Ariyawansa HA, Aoki T, Cardinali G, Crous PW, Druzhinina IS, Geiser DM,
753 Hawksworth DL, Hyde KD, Irinyi L, Jeewon R, Johnston PR, Kirk PM, Malosso E, May TW, Meyer W, Ôpik M, Robert
754 V, Stadler M, Thines M, Vu D, Yurkov AM, Zhang N, Schoch CL (2020) Unambiguous identification of fungi: where
755 do we stand and how accurate and precise is fungal DNA barcoding? *IMA Fungus* 11: doi:10.1186/s43008-020-
756 00033-z.
- 757 Lundberg DS, Yourstone S, Mieczkowski P, Jones CD, Dangl JL (2013) Practical innovations for high-throughput amplicon
758 sequencing. *Nature Methods* 2013 10:10 10: 999–1002. doi:10.1038/nmeth.2634.
- 759 Martin M (2011) Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnetjournal* 17: 10–12.
- 760 McMurdie PJ, Holmes S (2013) phyloseq: An R Package for Reproducible Interactive Analysis and Graphics of Microbiome
761 Census Data. *PLOS ONE* 8: doi:10.1371/journal.pone.0061217.
- 762 Nilsson RH, Anslan S, Bahram M, Wurzbacher C, Baldrian P, Tedersoo L, Nilsson RH, Anslan S, Bahram M, Wurzbacher C,
763 Baldrian P, Tedersoo L (2018) Mycobiome diversity: high-throughput sequencing and identification of fungi.
764 *Nature Reviews Microbiology* 2018 17:2 17: doi:10.1038/s41579-018-0116-y.
- 765 Ohta A, Nishi K, Hirota K, Matsuo Y (2023) Using nanopore sequencing to identify fungi from clinical samples with high
766 phylogenetic resolution. *Scientific Reports* 2023 13:1 13: 1–13. doi:10.1038/s41598-023-37016-0.
- 767 Özkurt E, Fritscher J, Soranzo N, Ng DYK, Davey RP, Bahram M, Hildebrand F, Özkurt E, Fritscher J, Soranzo N, Ng DYK, Davey
768 RP, Bahram M, Hildebrand F (2022) LotuS2: an ultrafast and highly accurate tool for amplicon sequencing analysis.
769 *Microbiome* 2022 10:1 10: doi:10.1186/s40168-022-01365-1.
- 770 Pastrana AM, Kirkpatrick SC, Kong M, Broome JC, Gordon TR (2017) *Fusarium oxysporum* f. sp. *mori*, a New Forma Specialis
771 Causing *Fusarium* Wilt of Blackberry. *Plant Disease* 101: doi:10.1094/PDIS-03-17-0428-RE.
- 772 Porras-Alfaro A, Liu K-L, Kuske CR, Xie G (2014) From Genus to Phylum: Large-Subunit and Internal Transcribed Spacer rRNA
773 Operon Regions Show Similar Classification Accuracies Influenced by Database Composition. *Applied and*
774 *Environmental Microbiology* 80: doi:10.1128/AEM.02894-13.
- 775 Rognes T, Flouri T, Nichols B, Quince C, Mahé F (2016) VSEARCH: a versatile open source tool for metagenomics. *PeerJ* 4:
776 e2584. doi:10.7717/peerj.2584.
- 777 Samson R, Visagie C, Houbraken J, Hong S-B, Hubka V, Klaassen C, Perrone G, Seifert K, Susca A, Tanney J, Varga J, Kocsubé
778 S, Szigeti G, Yaguchi T, Frisvad J (2014) Phylogeny, identification and nomenclature of the genus *Aspergillus*.
779 *Studies in Mycology* 78: doi:10.1016/j.simyco.2014.07.004.
- 780 Schoch CL, Seifert KA, Huhndorf S, Robert V, Spouge JL, Levesque CA, Chen W, Consortium FB, List FBCA, Bolchacova E,
781 Voigt K, Crous PW, Miller AN, Wingfield MJ, Aime MC, An K-D, Bai F-Y, Barreto RW, Begerow D, Bergeron M-J,
782 Blackwell M, Boekhout T, Bogale M, Boonyuen N, Burgaz AR, Buyck B, Cai L, Cai Q, Cardinali G, Chaverri P, Coppins
783 BJ, Crespo A, Cubas P, Cummings C, Damm U, Beer ZWd, Hoog GSd, Del-Prado R, Dentinger B, Diéguez-Uribeondo
784 J, Divakar PK, Douglas B, Dueñas M, Duong TA, Eberhardt U, Edwards JE, Elshahed MS, Fliegerova K, Furtado M,
785 García MA, Ge Z-W, Griffith GW, Griffiths K, Groenewald JZ, Groenewald M, Grube M, Gryzenhout M, Guo L-D,
786 Hagen F, Hambleton S, Hamelin RC, Hansen K, Harrold P, Heller G, Herrera C, Hirayama K, Hirooka Y, Ho H-M,

787 Hoffmann K, Hofstetter V, Högnabba F, Hollingsworth PM, Hong S-B, Hosaka K, Houbraken J, Hughes K, Huhtinen
788 S, Hyde KD, James T, Johnson EM, Johnson JE, Johnston PR, Jones EBG, Kelly LJ, Kirk PM, Knapp DG, Kõljalg U,
789 Kovács GM, Kurtzman CP, Landvik S, Leavitt SD, Liggenstoffer AS, Liimatainen K, Lombard L, Luangsa-ard JJ,
790 Lumbsch HT, Maganti H, Maharachchikumbura SSN, Martin MP, May TW, McTaggart AR, Methven AS, Meyer W,
791 Moncalvo J-M, Mongkolsamrit S, Nagy LG, Nilsson RH, Niskanen T, Nyilasi I, Okada G, Okane I, Olariaga I, Otte J,
792 Papp T, Park D, Petkovits T, Pino-Bodas R, Quaedvlieg W, Raja HA, Redecker D, Rintoul TL, Ruibal C, Sarmiento-
793 Ramírez JM, Schmitt I, Schüßler A, Shearer C, Sotome K, Stefani FOP, Stenroos S, Stielow B, Stockinger H, Suetrong
794 S, Suh S-O, Sung G-H, Suzuki M, Tanaka K, Tedersoo L, Telleria MT, Tretter E, Untereiner WA, Urbina H, Vágvölgyi
795 C, Vialle A, Vu TD, Walther G, Wang Q-M, Wang Y, Weir BS, Weiß M, White MM, Xu J, Yahr R, Yang ZL, Yurkov A,
796 Zamora J-C, Zhang N, Zhuang W-Y, Schindel D, Schoch CL, Seifert KA, Huhndorf S, Robert V, Spouge JL, Levesque
797 CA, Chen W, null n, null n, Bolchacova E, Voigt K, Crous PW, Miller AN, Wingfield MJ, Aime MC, An K-D, Bai F-Y,
798 Barreto RW, Begerow D, Bergeron M-J, Blackwell M, Boekhout T, Bogale M, Boonyuen N, Burgaz AR, Buyck B, Cai
799 L, Cai Q, Cardinali G, Chaverri P, Coppins BJ, Crespo A, Cubas P, Cummings C, Damm U, de Beer ZW, de Hoog GS,
800 Del-Prado R, Dentinger B, Diéguez-Uribeondo J, Divakar PK, Douglas B, Dueñas M, Duong TA, Eberhardt U,
801 Edwards JE, Elshahed MS, Fliegerova K, Furtado M, García MA, Ge Z-W, Griffith GW, Griffiths K, Groenewald JZ,
802 Groenewald M, Grube M, Gryzenhout M, Guo L-D, Hagen F, Hambleton S, Hamelin RC, Hansen K, Harrold P, Heller
803 G, Herrera C, Hirayama K, Hirooka Y, Ho H-M, Hoffmann K, Hofstetter V, Högnabba F, Hollingsworth PM, Hong S-B,
804 Hosaka K, Houbraken J, Hughes K, Huhtinen S, Hyde KD, James T, Johnson EM, Johnson JE, Johnston PR, Jones
805 EBG, Kelly LJ, Kirk PM, Knapp DG, Kõljalg U, Kovács GM, Kurtzman CP, Landvik S, Leavitt SD, Liggenstoffer AS,
806 Liimatainen K, Lombard L, Luangsa-ard JJ, Lumbsch HT, Maganti H, Maharachchikumbura SSN, Martin MP, May
807 TW, McTaggart AR, Methven AS, Meyer W, Moncalvo J-M, Mongkolsamrit S, Nagy LG, Nilsson RH, Niskanen T,
808 Nyilasi I, Okada G, Okane I, Olariaga I, Otte J, Papp T, Park D, Petkovits T, Pino-Bodas R, Quaedvlieg W, Raja HA,
809 Redecker D, Rintoul TL, Ruibal C, Sarmiento-Ramírez JM, Schmitt I, Schüßler A, Shearer C, Sotome K, Stefani FOP,
810 Stenroos S, Stielow B, Stockinger H, Suetrong S, Suh S-O, Sung G-H, Suzuki M, Tanaka K, Tedersoo L, Telleria MT,
811 Tretter E, Untereiner WA, Urbina H, Vágvölgyi C, Vialle A, Vu TD, Walther G, Wang Q-M, Wang Y, Weir BS, Weiß
812 M, White MM, Xu J, Yahr R, Yang ZL, Yurkov A, Zamora J-C, Zhang N, Zhuang W-Y, Schindel D (2012) Nuclear
813 ribosomal internal transcribed spacer (ITS) region as a universal DNA barcode marker for Fungi. *Proceedings of
814 the National Academy of Sciences* 109: doi:10.1073/pnas.1117018109.

815 Schubert K, Groenewald JZ, Braun U, Dijksterhuis J, Starink M, Hill CF, Zalar P, De Hoog GS, Crous PW (2007) Biodiversity in
816 the Cladosporium herbarum complex (Davidiellaceae, Capnodiales), with standardisation of methods for
817 Cladosporium taxonomy and diagnostics. *Studies in Mycology* 58: 105–156. doi:10.3114/sim.2007.58.05.

818 Shen W, Le S, Li Y, Hu F (2016) SeqKit: A Cross-Platform and Ultrafast Toolkit for FASTA/Q File Manipulation. *PLOS ONE* 11:
819 doi:10.1371/journal.pone.0163962.

820 Tedersoo L, Anslan S (2019) Towards PacBio-based pan-eukaryote metabarcoding using full-length ITS sequences.
821 *Environmental Microbiology Reports* 11: 659–668. doi:10.1111/1758-2229.12776.

822 Tedersoo L, Anslan S, Bahram M, Pöhlme S, Riit T, Liiv I, Kõljalg U, Kisand V, Nilsson H, Hildebrand F, Bork P, Abarenkov K,
823 Tedersoo L, Anslan S, Bahram M, Pöhlme S, Riit T, Liiv I, Kõljalg U, Kisand V, Nilsson H, Hildebrand F, Bork P,
824 Abarenkov K, Tedersoo L, Anslan S, Bahram M, Pöhlme S, Riit T, Liiv I, Kõljalg U, Kisand V, Nilsson H, Hildebrand F,
825 Bork P, Abarenkov K (2015) Shotgun metagenomes and multiple primer pair-barcode combinations of amplicons
826 reveal biases in metabarcoding analyses of fungi. *Mycology* 10: doi:10.3897/mycokeys.10.4852.

827 Tedersoo L, Bahram M, Zinger L, Nilsson RH, Kennedy PG, Yang T, Anslan S, Mikryukov V (2022) Best practices in
828 metabarcoding of fungi: From experimental design to results. *Molecular Ecology* 31: 2769–2795.
829 doi:10.1111/MEC.16460.

830 Tedersoo L, Hosseini Moghaddam MS, Mikryukov V, Hakimzadeh A, Bahram M, Nilsson RH, Yatsiuk I, Geisen S, Schwelm A,
831 Piwosz K, Prous M, Sildever S, Chmolewska D, Rueckert S, Skaloud P, Laas P, Tines M, Jung J-H, Choi JH, Alkahtani
832 S, Anslan S (2024) EUKARYOME: the rRNA gene reference database for identification of all eukaryotes. *Database*
833 2024: doi:10.1093/database/baae043.

834 Tedersoo L, Lindahl B (2016) Fungal identification biases in microbiome projects. *Environmental Microbiology Reports* 8:
835 774–779. doi:10.1111/1758-2229.12438.

836 Tedersoo L, Mikryukov V, Anslan S, Bahram M, Khalid AN, Corrales A, Agan A, Vasco-Palacios A-M, Saitta A, Antonelli A,
837 Rinaldi AC, Verbeken A, Sulistyo BP, Tamgnoue B, Furneaux B, Ritter CD, Nyamukondiwa C, Sharp C, Marín C, Dai
838 DQ, Gohar D, Sharmah D, Biersma EM, Cameron EK, De Crop E, Otsing E, Davydov EA, Albornoz FE, Brearley FQ,
839 Buegger F, Gates G, Zahn G, Bonito G, Hiiesalu I, Hiiesalu I, Zettur I, Barrio IC, Pärn J, Heilmann-Clausen J, Ankuda
840 J, Kupagme JY, Sarapuu J, Maciá-Vicente JG, Fovo JD, Geml J, Alatalo JM, Alvarez-Manjarrez J, Monkai J, Pöldmaa
841 K, Runnel K, Adamson K, Bräthen KA, Pritsch K, Tchan KI, Armolaitis K, Hyde KD, Newsham KK, Panksep K, Adebola
842 LA, Lamit LJ, Saba M, da Silva Cáceres ME, Tuomi M, Gryzenhout M, Bauters M, Bálint M, Wijayawardene N, Hagh-
843 Doust N, Yorou NS, Kurina O, Mortimer PE, Meidl P, Nilsson RH, Puusepp R, Casique-Valdés R, Drenkhan R,
844 Garibay-Orijel R, Godoy R, Alfarraj S, Rahimlou S, Pöhlme S, Dudov SV, Mundra S, Ahmed T, Netherway T, Henkel
845 TW, Roslin T, Fedosov VE, Onipchenko VG, Yasanthika WAE, Lim YW, Piepenbring M, Klavina D, Kõljalg U,
846 Abarenkov K, Tedersoo L, Mikryukov V, Anslan S, Bahram M, Khalid AN, Corrales A, Agan A, Vasco-Palacios A-M,
847 Saitta A, Antonelli A, Rinaldi AC, Verbeken A, Sulistyo BP, Tamgnoue B, Furneaux B, Ritter CD, Nyamukondiwa C,
848 Sharp C, Marín C, Dai DQ, Gohar D, Sharmah D, Biersma EM, Cameron EK, De Crop E, Otsing E, Davydov EA,
849 Albornoz FE, Brearley FQ, Buegger F, Gates G, Zahn G, Bonito G, Hiiesalu I, Hiiesalu I, Zettur I, Barrio IC, Pärn J,
850 Heilmann-Clausen J, Ankuda J, Kupagme JY, Sarapuu J, Maciá-Vicente JG, Fovo JD, Geml J, Alatalo JM, Alvarez-

- 851 Manjarrez J, Monkai J, Põldmaa K, Runnel K, Adamson K, Bråthen KA, Pritsch K, Tchan KI, Armolaitis K, Hyde KD,
852 Newsham KK, Panksep K, Adebola LA, Lamit LJ, Saba M, da Silva Cáceres ME, Tuomi M, Gryzenhout M, Bauters M,
853 Bálint M, Wijayawardene N, Hagh-Doust N, Yorou NS, Kurina O, Mortimer PE, Meidl P, Nilsson RH, Puusepp R,
854 Casique-Valdés R, Drenkhan R, Garibay-Orijel R, Godoy R, Alfarraj S, Rahimlou S, Põlme S, Dudov SV, Mundra S,
855 Ahmed T, Netherway T, Henkel TW, Roslin T, Fedosov VE, Onipchenko VG, Yasanthika WAE, Lim YW, Piepenbring
856 M, Klavina D, Kõljalg U, Abarenkov K (2021) The Global Soil Mycobiome consortium dataset for boosting fungal
857 diversity research. *Fungal Diversity* 2021 111:1 111: doi:10.1007/s13225-021-00493-7.
- 858 Tedersoo L, Tooming-Klunderud A, Anslan S (2018) PacBio metabarcoding of Fungi and other eukaryotes: errors, biases and
859 perspectives. *New Phytologist* 217: 1370–1385. doi:10.1111/nph.14776.
- 860 Vierstraete AR, Braeckman BP (2022) Amplicon_sorter: A tool for reference-free amplicon sorting based on sequence
861 similarity and for building consensus sequences. *Ecology and Evolution* 12: doi:10.1002/ece3.8603.
- 862 Vilgalys R, Hester M (1990) Rapid genetic identification and mapping of enzymatically amplified ribosomal DNA from several
863 *Cryptococcus* species. *Journal of Bacteriology* 172: 4238–4246. doi:10.1128/jb.172.8.4238-4246.1990.
- 864 Watanabe M, Yonezawa T, Lee K-i, Kumagai S, Sugita-Konishi Y, Goto K, Hara-Kudo Y (2011) Molecular phylogeny of the
865 higher and lower taxonomy of the *Fusarium* genus and differences in the evolutionary histories of multiple genes.
866 *BMC Evolutionary Biology* 11: doi:10.1186/1471-2148-11-322.
- 867 Went FW, Stark N (1968) THE BIOLOGICAL AND MECHANICAL ROLE OF SOIL FUNGI. *Proceedings of the National Academy of*
868 *Sciences* 60: 497–504. doi:10.1073/pnas.60.2.497.
- 869 Whitaker BK (2025) Design and Validation of a Peptide Nucleic Acid Clamp of Barley and Wheat ITS2 for Fungal Microbiome
870 Surveys. <https://doi.org/10.1094/PHYTOFR-07-25-0066-SC>: doi:10.1094/PHYTOFR-07-25-0066-SC.
- 871 White TJ, Bruns T, Lee S, Taylor J (1990) Amplification and direct sequencing of fungal ribosomal RNA genes for
872 phylogenetics. In: *PCR Protocols*. Elsevier, 315–322. doi:10.1016/b978-0-12-372180-8.50042-1.
- 873 Winand R, D’Hooge E, Van Uffelen A, Bogaerts B, Van Braekel J, Hoffman S, Roosens NHJ, Becker P, De Keersmaecker SCJ,
874 Vanneste K (2025) Investigating fungal diversity through metabarcoding for environmental samples: assessment
875 of ITS1 and ITS2 Illumina sequencing using multiple defined mock communities with different classification
876 methods and reference databases. *BMC Genomics* 26: doi:10.1186/s12864-025-11917-y.
- 877 Zeilinger S, Gupta VK, Dahms TES, Silva RN, Singh HB, Upadhyay RS, Gomes EV, Tsui CK-M, S CN (2015) Friends or foes?
878 Emerging insights from fungal interactions with plants. *FEMS Microbiology Reviews* 40:
879 doi:10.1093/femsre/fuv045.
- 880 Zhang T, Li H, Ma S, Cao J, Liao H, Huang Q, Chen W (2023) The newest Oxford Nanopore R10.4.1 full-length 16S rRNA
881 sequencing enables the accurate resolution of species-level microbial community profiling. *Applied and*
882 *Environmental Microbiology* 89: doi:10.1128/aem.00605-23.

883

884

885

886

887

888

889

890

891

892

893