Effective metabarcoding of endophytic fungi through host-exclusive primer design for nanopore long-read sequencing

Ting He[1], Koert Jansonius[1], Xiao Li[1], Alison M. Reilly[1], Bahar Sevgin[2], Rita Setroikromo[1], Thomas Hackl[2] and Kristina Haslinger[1,*]

1 Department of Chemical and Pharmaceutical Biology, Groningen Research Institute of Pharmacy, University of Groningen, Antonius Deusinglaan 1, 9713 AV Groningen, The Netherlands; t.he@rug.nl (T.H.); k.jansonius@student.rug.nl (K.J.); xiao.li@rug.nl (X.L.); a.m.reilly@student.rug.nl (A.M.R): r.setroikromo@rug.nl (R.S.)

2 Groningen Institute for Evolutionary Life Sciences, University of Groningen, Nijenborgh 7, 9747 AG Groningen, The Netherlands; t.hackl@rug.nl; b.sevgin@student.rug.nl (B.S.)

*Correspondence: k.haslinger@rug.nl

Introduction: 670 words
Material and Methods: 1469 words
Results: 2710 words
Discussion: 873 words
5 color figures
4 tables
Supporting Information: one pdf with supporting figures S1-S3, supporting tables S1-2, and note S1; one tabulated data file with supporting data 1-8.

Summary

- Metabarcoding is a powerful tool to simultaneously identify multiple taxa within a DNA sample. However, despite its power, metabarcoding poses challenges when applied to host-associated microbiomes, primarily due to the substantial co-amplification of host DNA.
- Here we developed a host-exclusive primer design workflow, to selectively generate amplicons from target taxa while excluding the host. This workflow is centered around a new computational tool, mbc-prime, that can generate a list of discriminative candidate primers and score them. We showcase the use of this tool in the process of designing primers for the long-read metabarcoding of endophytic fungi in *Vinca minor.*
- Four new forward and two new reverse primers reliably amplify large regions of the fungal rDNA of a fungal mock community and the extracted holobiont genomic DNA. Primers exhibit high specificity for fungal rDNA and robust amplification across a diverse range of fungal taxa: with the best primer pairs, we produced fungal amplicons free of plant DNA.
- Our workflow can be used to study the composition of complex host-associated microbiomes. This workflow should furthermore be universally applicable for the design of discriminative primers in a user-friendly and practical manner and thus be of use for various researchers in microbiome research.

Keywords

Amplicon sequencing, fungal taxonomy, host-associated microbiome, ITS, Oxford Nanopore Technology, plant microbiome, primer design, 28S, 18S

Introduction

High-throughput sequencing of taxonomic marker genes (metabarcoding) is a widely used technique for biodiversity profiling and microbial ecology (Abdelfattah *et al.*, 2018). The choice of primers is the most important step for metabarcoding in order to achieve an appropriate coverage of the species in the target phyla and potentially exclude non-target phyla (Tedersoo *et al.*, 2022). Although different primer design tools, such as primer3 (Rozen & Skaletsky, 2000), Primer BLAST (Ye *et al.*, 2012) and PrimerQuest™ (https://www.idtdna.com/PrimerQuest), are designed to suit various primer design needs, they have some limitations that have hindered their widespread adoption in ecology: they do not allow for the design of discriminative primers that selectively amplify regions in the genomes of target phyla but not others. Such discriminative power is typically necessary when working with complex samples that cannot be enriched for the target phylum prior to or during DNA extraction. One such example are host-

53 associated microbiomes, where bulk DNA extractions often include a large amount of host DNA (Yap
54 *et al.*, 2020).

55 Host-associated microbiomes in general are important because they support health and resilience of the
56 host. For example, plants live in close association with fungi contributing significantly to nutrient
57 cycling in the soil (Parniske, 2008), pathogen defense (Bolívar-Anillo *et al.*, 2020), and stress tolerance
58 (Rodriguez *et al.*, 2008). The taxonomic characterization of fungal communities within plant
59 microbiomes is crucial for understanding their roles in plant-microbe interactions and ecosystem
60 functioning (Dastogeer *et al.*, 2020). Most studies targeting fungi employ the internal transcribed spacers
61 1 and 2 (ITS1 and ITS2) region of ribosomal DNA, which was proposed as the official primary barcode
62 for fungi (Schoch *et al.*, 2012). However, the widely used ITS primers, such as ITS1 and ITS4 (White
63 *et al.*, 1990), as well as ITS1F (Gardes et al., 1993), designed to target a broad range of fungi, cannot
64 exclude amplification of the plant ITS, which is often highly abundant in typical samples can account
65 for more than 90% of the data (Toju *et al.*, 2012; Ihrmark *et al.*, 2012).

66 Furthermore, environmental sequencing studies based on ITS sequences alone regularly recover fungal
67 sequences that are difficult to assign to any fungal lineage (Rosling *et al.*, 2011) and feature a significant
68 proportion of sequences classified as "unidentified fungi" due to the lack of more explicit taxonomic
69 information (Tedersoo *et al.*, 2014; Heeger *et al.*, 2019; Ceballos-Escalera *et al.*, 2022; Monod *et al.*,
70 2022). Examining more conserved, higher order taxonomic markers, like the small ribosomal subunit
71 (SSU) and large ribosomal subunit (LSU) can help resolve such uncertainties at the expense of
72 taxonomic resolution at the species level (Raja *et al.*, 2017). There is a growing interest in applying
73 long-read sequencing of amplicons covering the entire rDNA locus to simultaneously benefit from the
74 higher order information from SSU and LSU, and the species-level resolution provided by the ITS when
75 studying taxonomically complex groups (De Filippis *et al.*, 2017; Wurzbacher *et al.*, 2019; Vu *et al.*,
76 2019). In light of the recent improvements in read accuracy of the Oxford Nanopore Technology, which
77 reaches now more than 99.0% accuracy (https://nanoporetech.com/accuracy), long-read sequencing is
78 now offering a competitive alternative to short-read amplicon sequencing (Cuscó *et al.*, 2018; Tedersoo
79 *et al.*, 2021; Charnaud *et al.*, 2022).

80 However, to our knowledge, no extensive evaluation of fungus-specific primers for rDNA locus
81 sequencing in the plant microbiome has been established so far. Thus far, there has been only one
82 reported primer (LR5-Fung) identified to specifically target the LSU gene, showing a preference for
83 amplifying the fungal LSU over that of plants (Tedersoo *et al.*, 2008). This highlights the need for
84 developing new fungus-specific primers that target both the SSU and LSU genes, with the aim of
85 excluding plant DNA contamination.

86 Here, we describe the development of an effective metabarcoding workflow for host-associated
87 microbiomes, starting with the design of discriminative primers with our new *mbc-prime* tool that
88 detects and ranks potential discriminative primers. We demonstrate the execution of this workflow by
89 targeting the community of endophytic fungi residing in the evergreen plant lesser periwinkle (*Vinca*
90 *minor*).

91 Materials and Methods

92 Sample collection and DNA extraction from the leaves of *Vinca minor*

93 For the sampling, 10 - 15 individual plants of *Vinca minor* were collected from Italy (45°26'3'' N,
94 8°49'53''W), immediately placed in an ice box, and transported to the lab. After surface sterilization
95 (He *et al.*, 2023), the plant material was quickly frozen with liquid nitrogen, ground into fine powder,
96 and stored at – 70 °C until DNA extraction.

97 Holobiont genomic DNA was extracted using NucleoSpin® Microbial DNA kit (BIOKÉ, Leiden, The
98 Netherlands) in accordance with the protocols provided by the manufacturer with adjustment. Briefly,
99 0.2 g ground plant material was suspended in 100 µL elution buffer BE, 40 µL buffer MG and 10 µL
100 liquid proteinase K. The plant cells were disrupted by vortexing at 1000 rpm for 3 mins. After adding
101 600 µL buffer MG, samples were vortexed for another 1 min at 1000 rpm, followed by 2 min
102 centrifugation at 13,000 x g. The supernatant was transferred onto a NucleoSpin® Microbial column,
103 washed twice with buffer BW and B5 buffer, and eluted with 30 µL elution buffer BE. The DNA was
104 quantified with NanoDrop N-100 (ThermoFisher, Waltham, MA, USA) and stored in – 20 °C until
105 further use.

106 Mock community construction

107 To generate a synthetic mock community for initial testing of the new primers, we used six species
108 isolated from healthy-looking, surface-sterilized leaves of *Vinca minor,* collected in Groningen (The
109 Netherlands) in November 2021. Fungal isolation, cultivation, and genomic DNA isolation were
110 executed as described in our previous work (He *et al.*, 2023). The SR1R-Fw/ LR12-R primer pair
111 (D'Andreano *et al.*, 2021) was used to amplify the entire fungal rDNA locus. Reactions were run in
112 volumes of 25 µL, containing 1 µL each of the forward (10 µM) and reverse primer (10 µM), 12.5 µL
113 LongAmp Hot Start Taq 2x Master Mix (New England Biolab, MA, USA), 1 µL of DNA template (40
114 fmol), and 9.5 µL of nuclease-free water. PCR amplification was carried out with the following protocol:
115 initial denaturation of 94°C for 30 s, followed by 30 cycles of 94°C for 10 s, 65°C for 30 s, and 65°C
116 for 5.5 min, followed by a final 65°C extension for 10 min. The PCR products were purified using
117 AMPure XP magnetic beads (Beckman Coulter, California, USA) and the concentrations were
118 determined using Qubit 3.0 (Invitrogen, Waltham, MA, USA), and the Qubit dsDNA HS Assay Kit.

119 These six fungal amplicons were mixed in equimolar ratios to construct the mock community DNA
120 sample with a concentration of 55 ng/ µL.

121 Primer design and implementation on endophytic fungi in *Vinca minor*

122 To analyze endophytic fungi in *Vinca minor*, all related sequences were downloaded from the SILVA
123 database (Datasets S1 and S2). For the SSU region, 46 sequences from the Gentianales order were
124 extracted from the downloaded database file (SILVA_138.1_SSURef_NR99_tax_silva.fasta.gz.). In the
125 same way, a dataset of 118 fungal sequences was collected, which represents all major fungal divisions
126 (ascomycota, basidiomycota, chytridiomycota, cryptomycota, zoopagomycota, microsporidia,
127 mucoromycota, neocallimastigomycota and blastocladiomycota). These 164 sequences were aligned
128 using the MAFFT multiple sequence alignment software (ouput format: fasta format, input order;
129 strategy: L-INS-i; arguments: --ep)(Katoh & Standley, 2013). For the LSU region, 62 fungal sequences
130 (Dataset S3) and 6 available Gentianales sequences (Dataset S4) were extracted from the downloaded
131 database file (SILVA_138.1_LSURef_NR99_tax_silva.fasta.gz). Alignment was performed among
132 these 68 sequences in the same way as for the SSU.

133 Both alignments were analyzed with mbc-prime version 0.6.0 (-s = 0.5) (see Results section for
134 implementation). We obtained 147 candidate discriminative primers across the SSU (Dataset S5) and
135 371 across the LSU (Dataset S6). Four candidates for forward primers on the SSU and two reverse
136 primers on the LSU were selected from the initial set based on the following criteria: 1) at least 80%
137 coverage of fungi at zero or one mismatch (*in silico*); 2) at least one mismatch at the 3' end between
138 primer compared to the plant sequences; 3) a GC content between 40 and 60%; 4) approx. 20 base pairs
139 in length. In addition, we manually optimized the length of candidate primer sequences to lower the
140 difference in melting temperatures and the likelihood of heterodimer formation in primer combinations
141 as well as incorporated degenerate positions to maximize the coverage of the target group. Finally, all
142 selected primers were evaluated *in silico* using SILVA TestProbe 3.0 (https://www.arb-
143 silva.de/search/testprobe/) for the individual primers and OligoAnalyzer™ (Version 3.1, Integrated
144 DNA Technologies, https://eu.idtdna.com/pages/tools/oligoanalyzer) for primer pairs before being
145 tested on the DNA samples of a fungal mock community and plant material. A step-by-step tutorial for
146 using mbc-prime is listed in the Note S1.

147 Amplicon generation for Oxford Nanopore sequencing

148 The strategy employed in this study involves two steps of PCR. The first PCR was performed with the
149 gene-specific rDNA primers including overhangs that serve as handles for PCR barcoding in the second
150 step (5' TTTCTGTTGGTGCTGATATTGC-[forward primer sequence] 3'; 5'
151 ACTTGCCTGTCGCTCTATCTTC-[reverse primer sequence] 3'). All primer pairs used in this study
152 are shown in Table 2 and sequences of all primers used in this study are listed in Dataset S7. Each PCR

153 reaction mix consisted of 0.4 µM each of the forward and reverse primer, 1 x LongAmp Hot Start Taq
154 2x Master Mix (New England Biolab, MA, USA), and 40 ng of template DNA, combined in a 25 µL
155 reaction. The PCR cycling conditions were as follows: initial denaturation of 94°C for 30 s, followed by
156 30 cycles of a 10 s hold at 94°C, a 30 s hold at the respective annealing temperature shown in Table 1,
157 and an elongation period specific for the length of the amplicon as shown in table 1 at 65°C, followed
158 by a final 65°C extension for 10 min. The PCR amplicons were purified using AMPure XP magnetic
159 beads (Beckman Coulter, California, USA) and quantified by Qubit. The second PCR was used to
160 incorporate the Oxford Nanopore barcode sequences with the PCR Barcoding Expansion 1-12 (EXP-
161 PBC001, ONT, Oxford, UK). Each amplicon PCR product was used as a template (100-200 fmol) in
162 the second PCR reaction with 1 µL PCR barcoding primer mix (one of BC1- BC12, at 10 µM), and 25
163 µL LongAmp Hot Start Taq 2X Master Mix (New England Biolab, MA, USA) in a total of 50 µL. The
164 PCR cycling conditions were as follows: initial denaturation of 95°C for 3 min, followed by 15 cycles
165 of 95°C for 15 s, 62°C for 15 s, and an elongation period specific for the length of the amplicon as shown
166 in table 1 at 65°C, followed by a final 65°C extension for 10 min. The barcoded PCR products were
167 further purified with 0.4x volume of AMPure XP Beads (Beckman Coulter, California, USA). All
168 barcoded PCR products were pooled in equal amounts to prepare 1 µg of DNA in 50 µL nuclease-free
169 water for library preparation. The metadata of samples as input for the library preparation is listed in the
170 Dataset S8.

171 Library preparation

172 The library preparation was performed with the ligation sequencing kit V14 (ONT, SQK-LSK114)
173 following the manufacturer's instructions
174 (https://community.nanoporetech.com/docs/prepare/library_prep_protocols/ligation-sequencing-gdna-
175 pcr-barcoding-sqk-lsk114-with-exp/v/pbc_9182_v114_revh_07mar2023). First, the pooled barcoded
176 amplicons underwent end repair with the NEBNext® Ultra II End Repair / dA-tailing Module (NEB,
177 E7546). This reaction consisted of 50 µL DNA (1 µg), 7 µL Ultra II End-prep Reaction Buffer and 3
178 µL Ultra II End-prep Enzyme Mix, and was incubated at 20°C for 5 minutes and 65°C for 5 minutes.
179 The end-repaired amplicons were further purified using AMPure XP magnetic beads and quantified by
180 Qubit. For the adapter ligation step, 60 µL of the end-repaired and purified amplicons, 25 µL Ligation
181 Buffer (LNB), 10 µL NEBNext Quick T4 DNA Ligase, and 5 µL Ligation Adapter (LA) were combined
182 in a reaction tube and incubated for 10 min at room temperature. After ligation, the samples were mixed
183 with 0.4 x volume of AMPure XP magnetic beads and then washed twice with 250 µL Long Fragment
184 Buffer provided in the ligation sequencing kit. The library was finally eluted with 15 µL Elution Buffer
185 (EB), incubated at 37 °C for 10 minutes, and quantified by Qubit. 35-50 fmol of this final prepared
186 library were loaded for sequencing.MinION sequencing data analysis

Sequencing was performed on a MinION device with a R10.4.1 flow cell running with MinKNOW software v24.02.10 . Basecalling and demultiplexing were performed with Dorado v7.3.9 in super accurate mode. We trimmed reads using NanoFilt v2.8.0 and assessed the read quality using NanoPlot v1.42.0. Subsequently, the FASTQ files were uploaded to EPI2ME Labs v4.1.4, and the "wf-metagenomics" workflow was performed with the following settings: classifier: Kraken2; Kraken2 memory mapping: true; dataset: ncbi_16s_18s_28s_ITS or SILVA_138_1; N taxa barplot: 9; bracken length: 2000; min percent identity: 90; min len: 2000; max len: 6000. Relative abundances, Alpha diversity index (richness, Simpson and Shannon) and species richness curves were generated as part of the analysis workflow and further visualized with GraphPad Prism 10.

Results

A workflow for effective host-exclusive primer design

Sensitive metabarcoding primers able to amplify marker regions equally well across a broad taxonomic range require a pair of conserved sequence loci surrounding a variable region with enough differences to distinguish individual species. This pattern, for example, is found in the commonly used primers for the 16S gene variable regions or the highly variable rRNA internal transcribed spacers (ITS/ITS2), which are surrounded by the conserved small 18S and the large 28S rRNA subunits (Klindworth *et al.*, 2013).

Expanding on this pattern, the rationale for creating host-exclusive, discriminative primers is simple: identify a variable-region-enclosing primer pair that is conserved among all target group members but differs in sequence in all unwanted taxa. Such locations can be discerned in one or more combined multiple-sequence alignments (MSAs) of representative sequences of the target and exclusion group. This rationale allowed us to devise an effective workflow for host-exclusive primer design for holobiont metabarcoding (Figure 1). Since this workflow requires identifying discriminative primer sites in MSAs, which is slow and tedious if done by hand, we also devised the program mbc-prime (described in detail in the next section) to help automate this crucial step. The list of discriminative, candidate primers returned by mbc-prime can then be further inspected for desired properties of the primer, for instance the GC content and melting temperature of the primer and the position of the non-target mismatches within the primer sequence. With the most discriminative primers in hand, additional optimization of the primers can be performed by introducing degenerate bases into the sequence to optimize the coverage among the target sequences. Lastly, the candidate primers can be evaluated for their specificity for target group and efficiency of excluding non-target group.

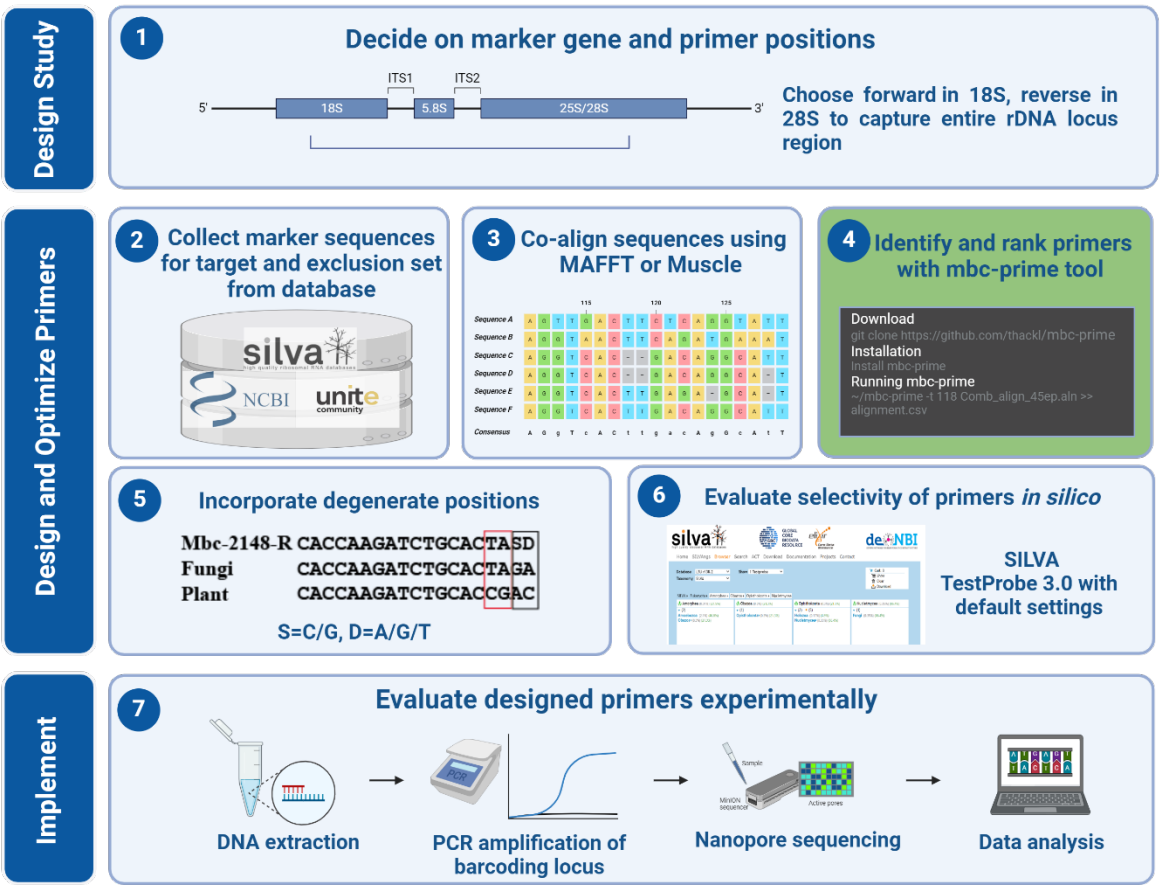

Figure 1. Workflow for effective host-exclusive primer design as described in this study. At the core of the workflow stands the computational tool mbc-prime, which effectively identifies candidates for discriminative primers based on multiple sequence alignments of the target and the exclusion sets (step 4).

Detection of discriminative primers with mbc-prime

Discriminative primers should align optimally with all target sequences while differing maximally from all non-target sequences. To identify such primers computationally, we developed mbc-prime, a stand-alone Python program available at https://github.com/thackl/mbc-prime.

mbc-prime identifies discriminative primers based on a user-supplied MSA in FASTA format, comprising DNA sequences of the chosen genomic locus in the target and exclusion sets. In a first step, the MSA is imported and preprocessed. Terminal gaps, introduced by partial sequences frequently present in amplicon marker databases, are masked to distinguish them from genuine gaps. Columns with more than 20% (default value) gaps or ambiguous bases (non-ATGCU) are trimmed from the alignment. The trimmed MSA is then analyzed via a sliding window with a size corresponding to the intended

234 primer length (default: 20 bp). For each window and sequence set (target and exclusion), the abundance
235 of each sequence variant is counted, and the most abundant sequence variant of the target set is chosen
236 as primer candidate. Next, for each remaining sequence variant from both sets, the number of
237 mismatches compared to the primer candidate is computed, and the overall percentage of sequence
238 variants with 0, 1, 2, 3 or 4-or-more mismatches are determined. The underlying edit distances are
239 computed via Python bindings of the Edlib library (Šošić & Šikić, 2017). This results in two mismatch
240 profiles per primer candidate, one for the target set and one for the exclusion set. An example of
241 promising profiles for a discriminative primer would be the following: 90% and 10% of the target
242 sequences align with 0 and 1 mismatch, respectively, while 10%, 20% and 70% of the exclusion
243 sequences align with 2, 3, and 4-or-more mismatches, respectively.

244 To facilitate easy comparison and ranking of possible primer candidates based on their mismatch profiles,
245 we integrate the two profiles obtained from target and exclusions sets into a single discriminative score,
246 as defined below. The score ranges between 0 and 1, with 1 capturing 100% of target sequences perfectly
247 while excluding 100% of non-target sequences at the given mismatch threshold (default mismatch
248 threshold: mm = 2).

$$\text{score} = \frac{1}{mm} \sum_{k=1}^{mm} \left(\sum_{i=1}^{k} \text{target}[i] - \sum_{i=1}^{k} \text{exclude}[i] \right)$$
249

250 Finally, candidate primers in the MSA are filtered for a maximum number of gaps, ambiguous bases
251 and a minimum score (default 0.5), and overlapping primers are grouped together. The filtered list of
252 above-threshold primer candidates is reported together with the positional information, score, mismatch
253 profiles, and sequence in a tab-separated table.

254 Design of metabarcoding primers for endophytic fungi in *Vinca minor*

255 To demonstrate the applicability of our metabarcoding primer design workflow, we targeted the
256 endophytic mycobiome of *Vinca minor* (lesser periwinkle) with long-read sequencing of the rDNA locus.
257 *V. minor* belongs to the order of Gentianales in the Apocynaceae family and is a popular plant for its
258 ornamental and medicinal properties. It produces a wide array of vinca alkaloids with neuroprotective
259 and antioxidant bioactivities and was used in folk medicine for the treatment of hypertension, as a
260 carminative, emetic, hemostatic, and astringent, and in the treatment of toothache and snakebite (Vas &
261 Gulyás, 2005; Bahadori *et al.*, 2012). More than 50 indole alkaloids have been extracted and isolated
262 from *V. minor* (Farahanikia *et al.*, 2011). Its main alkaloid, vincamine is a recognized neuroprotector
263 and cerebral vasodilator, widely used in the pharmaceutical industry (Fandy *et al.*, 2016). To date, the
264 microbiome of *V. minor* remains largely unexplored. In our initial amplicon sequencing experiments
265 with published primers targeting the rDNA of fungi for long-read sequencing, we observed that 99 %

of reads were derived from the *V. minor* rDNA locus (data not shown). Therefore, we considered this the ideal test case for the performance of our holobiont metabarcoding workflow.

Since we wanted to target the entire rDNA locus, we downloaded fungal (target) and plant (exclusion) sequences spanning the SSU and LSU regions from the SILVA database. We selected 118 fungal SSU sequences from different phyla (Ascomycota, Basidiomycota, Chytridiomycota, Cryptomycota, Zoopagomycota, Microsporidia, Mucoromycota, Neocallimastigomycota, and Blastocladiomycota) to represent all major fungal divisions (Dataset S1), and 46 sequences from the order Gentianales (Dataset S2). For the LSU, we included 62 fungal sequences (Dataset S3) and 6 sequences from Gentianales (Dataset S4). We created two MSAs for all SSU and LSU sequences, respectively, and scanned them for discriminative primer binding sites. Using mbc-prime with default settings, we obtained 147 candidate primers across the SSU (Dataset S5) and 371 across the LSU (Dataset S6). From the list, we selected primers that had scores >0.6 and carried at least one mismatch at the 3' end when compared to plant (exclusion) sequences. With these criteria we obtained four forward primers in different regions of the SSU and two reverse primers in different regions of the LSU (Figure 2A). Following careful examination of the GC content, the predicted melting temperatures, and the respective primer binding sites within the MSA, we manually optimized the primer sequences to establish the final sequences (Figure 2B, Dataset S7).

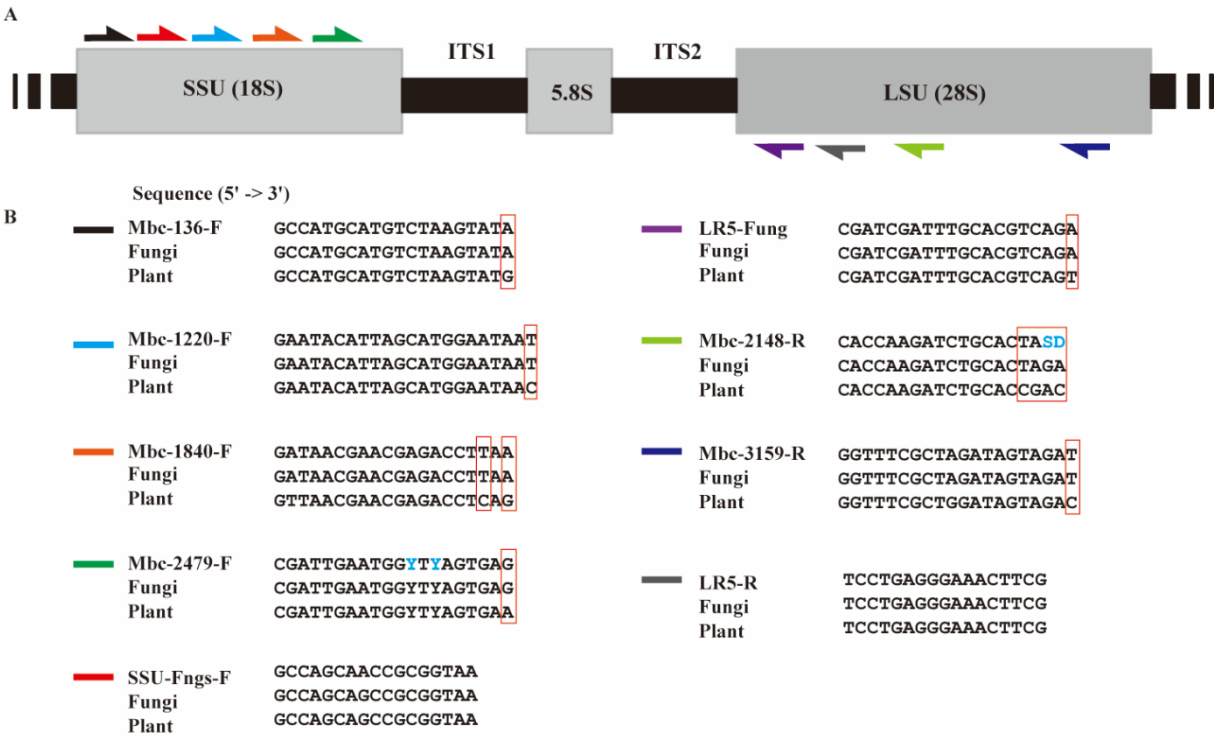

Figure 2. A. Illustration of primer locations in the rDNA locus consisting of SSU, ITS1, 5.8S, ITS2, and LSU; B. Sequence alignments of primers with the consensus sequences of fungi (target set) and plant (exclusion set). Degenerate bases (marked in blue): Y = C/T, S = C/G, D = A/G/T.

287 Primer mbc-1220-F was extended by two nucleobases at the ‘5 end, mbc-2148-R was shortened by one
288 base at the ‘5 end, and in mbc-2479-F and mbc-2148-R we incorporated degenerate positions to
289 maximize the coverage of the target group (Fig. S1).

290 To assess the performance of the final primers *in silico*, we used the SILVA TestProbe 3.0 tool with
291 default settings and “fungi” and “gentianales” as test sets. For completeness, we also included several
292 primers from literature (Tedersoo et al., 2008, 2015;Nilsson et al., 2019). Similar to the literature primers,
293 all of the new primers show a wide coverage across fungi. In addition, several literature primers show
294 coverage within the Gentianales, while our new primers show almost no coverage with either zero or
295 one mismatch (Table 1). The only literature primer with the desired specificity and a high mbc-prime
296 weighted score is the LR5-Fung reverse primer.

297 With these primer candidates in hand, we performed an *in silico* analysis of all possible primer
298 combinations to identify those with favorable parameters, such as a low difference in melting
299 temperatures and a low likelihood of heterodimer formation. The 10 primer pairs that met our quality
300 standards (pairs 1-10; Table 2, Table S1) and one primer pair from literature (“3.5 kb”) (D’Andreano *et*
301 *al.*, 2021) were carried forward for experimental validation with a fungal mock community and genomic
302 DNA extracted from the surface-sterilized *V. minor* holobiont.

303 Experimental evaluation of the designed fungus-specific primers using mock 304 community

305 We generated a standardized DNA sample by mixing equimolar amounts of PCR-amplified rDNA from
306 six fungal isolates of our in-house strain collection (Table S2). The strains were chosen to cover a wide
307 taxonomic range and different levels of GC content in order to assemble an authentic mock community
308 for our primer testing. Next, we optimized the annealing temperatures of the PCR for each primer pair
309 to generate high-quality amplicons based on visual inspection of the bands after agarose gel

310 **Table 1:** mbc-prime and SILVA TestProbe 3.0 evaluations of all primers used in this study.

Primer	Weighted score[a]	Coverage of Fungi[b]			Coverage of Gentianales[b]			Source
		0 mm	1 mm	2 mm	0 mm	1 mm	2 mm	
SSU-Fngs-F	-	0.2%	94.8%	96.4%	0%	80.0%	88.9%	Tedersoo et al. 2015
mbc-2479-F	0.62	77.1%	87.8%	89.2%	0%	0%	0%	This study
mbc-1840-F	0.62	66.1%	83.5%	90.2%	0%	0%	2.2%	This study
mbc-136-F	0.78	80.1%	89.7%	92.3%	0%	0%	87.5%	This study
mbc-1220-F	0.66	52.1%	82.3%	88.7%	0%	0%	11.1%	This study
LR5-Fung	0.75	88.0%	91.8%	93.3%	0%	0%	100.0%	Tedersoo et al. 2008
mbc-3159-R	0.80	87.0%	92.5%	95.1%	0%	0%	100.0%	This study
mbc-2148-R	0.72	86.4%	92.0%	94.4%	0%	0%	0%	This study
LR5-R	-	90.9%	90.3%	94.2%	100.0%	100.0%	100.0%	Tedersoo et al. 2015

311 a: mbc-prime weighted score of the candidate primer before manual modification (optional); b: summative
312 coverage achieved with the final primers determined by SILVA TestProbe 3.0; mm: number of maximum
313 mismatches

314 **Table 2** All primer pairs used in this study.

Pair	Primers (fwd/rev)	Tm (°C)	Product size
1	mbc-2479-F/LR5-Fung	50	1.4-1.8 kb
2	mbc-2479-F/mbc-3159-R	50	2.9-3.3 kb
3	mbc-2479-F/mbc-2148-R	49	2.1-2.4 kb
4	mbc-1840-F/mbc-3159-R	49	3.2-3.5 kb
5	mbc-1840-F/mbc-2148-R	50	2.2-2.7 kb
6	ITS1catta/LR5-Fung	55	1.3-1.6 kb
7	mbc-136-F/LR5-Fung	50	3.1-3.4 kb
8	mbc-136-F/mbc-3159-R	49	4.3-4.9 kb
9	mbc-136-F/mbc-2148-R	50	3.6-4.0 kb
10	mbc--1220-R/mbc-3159-R	46	3.7-5.0 kb
3.5kb	SSU-Fngs-F/LR5-R	55	3.5 kb

315 Tm: The optimized annealing temperatures are based on the LongAmp hot start polymerase

316 electrophoresis (Figure 3A). All primer pairs yielded clear bands of the appropriate molecular weight.
317 Since it was our primary goal to generate long amplicons covering both SSU and LSU, we abandoned
318 primer pairs 1-6 after this step and proceeded with multiplex sequencing of the amplicons generated
319 with primer pairs 7-10 with the R10.4.1 flow cells from Oxford Nanopore Technologies. All steps from
320 PCR to sequencing were performed with technical duplicates and in parallel to a positive control (3.5
321 kB primer pair) and a negative control (nuclease-free water). With one sequencing run, we generated
322 303,100 raw reads with a yield of 1.1 Gb and a mean Qscore of 18 (accuracy: 98%; Table 3).

323 We used the wf-metagenomics workflow with Kraken2 to classify the reads against the NCBI RefSeq
324 Targeted Loci database (ncbi_16s_18s_28s_ITS). Based on the ratio between classified reads and the
325 total reads, we calculated that for all samples more than 99.9% of reads were assigned to a taxonomic

326 **Table 3** Read statistics after de-multiplexing of sequenced amplicons generated from the mock
327 community with various primer pairs.

Primer set	Number of reads after basecalling	Number of reads after trimming*	Mean length (bp)	Mean Qscore	Total yield (Mb)
7	24,963	19,887	3,124	18.1	62.1
	35,782	29,996	3,125	18.2	93.7
8	19,008	11,776	4,498	18.0	53.0
	33,468	26,016	4,537	18.1	117.0
9	32,723	26,531	3,659	18.1	97.1
	39,945	33,761	3,644	18.3	123.0
10	30,433	22,166	3,812	18.1	84.5
	23,057	18,735	3,825	18.1	71.7
3.5 kb	27,129	18,767	2,745	17.9	51.5
	20,945	16,899	2,743	17.9	46.4

328 * Reads were trimmed in the range of 2 kb – 6 kb.

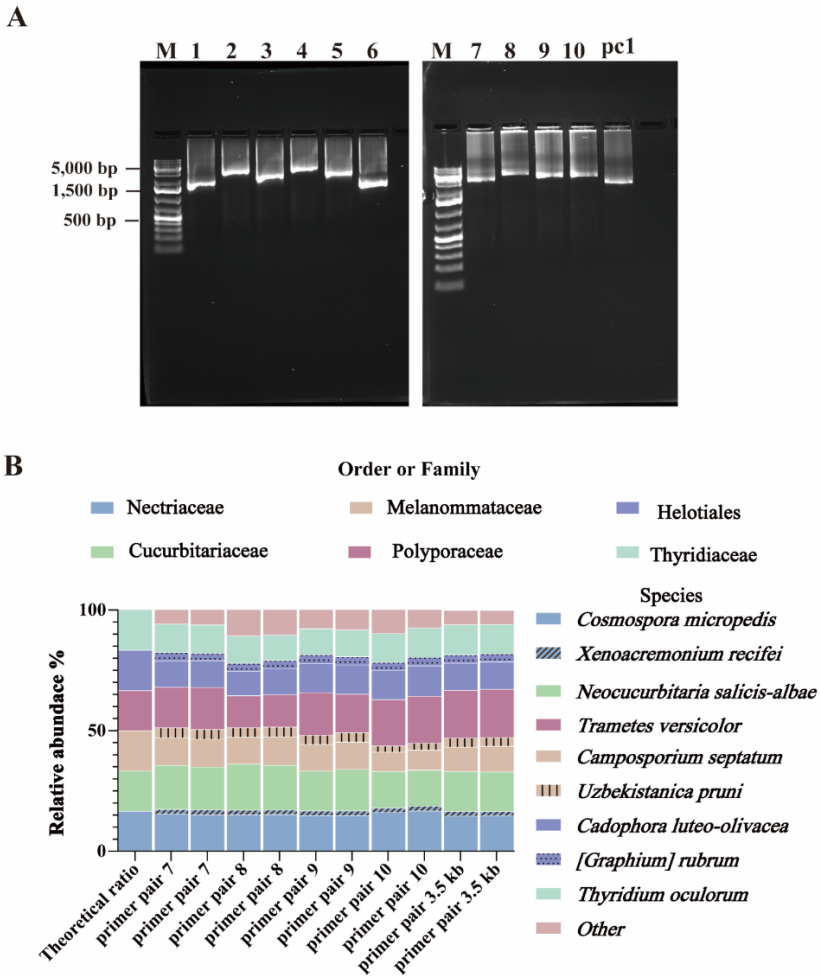

Figure 3. Metabarcoding of a fungal mock community. (A) Gel electrophoresis of rDNA amplicons generated from gene specific primers including overhangs for PCR barcoding; Sample order: M: 1 kb plus DNA ladder; 1-10: amplicons generated with primer pairs 1-10 respectively; pc1: positive control generated with primer pair 3.5 kb. All primer pairs yielded products in the expected size range. (B) Relative abundance of taxonomic groups classified based on the sequencing reads of amplicons generated from the mock community with the different primer pairs (each primer pair has two replicates). Species belonging to the same family are displayed in the same color with different fill patterns. All primer pairs reproduce the overall structure of the mock community, albeit with different biases for different groups.

group. Among the six isolates, four were correctly identified at species level (Figure 3B, Table S2). The isolates of *Aposphaeria corallinolutea* and *Cosmosporella olivacea*, which lack species representatives in the RefSeq database, were primarily assigned to two close relatives in the database, *Camposporium septatum* and *Cosmospora micropedis*, respectively. *Cosmosporella* and *Cosmospora* are genera in the family *Nectriaceae* (Huang *et al.*, 2018), while *Aposphaeria* and *Camposporium* are members of the Pleosporales (Hyde *et al.*, 2020). A minor fraction of the reads was moreover classified as *Uzbekistanica pruni*, which also belongs to the family of *Melanommataceae* and order of Pleosporales.

345 Overall, the new primers perform equally well in amplifying the rDNA of various fungal isolates as the
346 primer pair from literature, with primer sets 7 and 9 giving the best reflection of the mock community
347 composition. For completeness, we also repeated our classification analysis with the SILVA database.
348 We found that the results from the SILVA database did not closely align with the microbial compositions
349 in the mock community, with only the genus *Trametes* being accurately identified at the genus level
350 (Fig. S2). Therefore, we conclude that the analysis with NCBI is superior for the analysis of fungal
351 communities.

352 Experimental evaluation of the designed fungus-specific primers using genomic 353 holobiont DNA of *V. minor*

354 Having established our approach for four long-amplicon primer pairs in a mock community setting, we
355 next sought to assess the performance of the new fungi-specific primer pairs in a real-world holobiont
356 scenario. We applied these primer pairs to specifically target the endophytic fungi in the holobiont DNA
357 from *V. minor*. However, while preparing for the experiment and performing the first amplification step
358 with the holobiont DNA (Figure 4A), we noticed that primer pair 8 did not give a high-quality PCR
359 product with a distinct band of the correct size. The individual primers performed well in the other
360 primer combinations thus we conclude that the issue specifically arises from the combination of these
361 primers in the context of the holobiont sample, where the target fungal DNA is rare. Since we could not
362 overcome this problem despite various optimization attempts (annealing temperature, additives,
363 polymerase), we abandoned primer set 8.

364 Proceeding with the remaining three discriminative primer pairs 7, 9, and 10, and the non-discriminative
365 literature-based primer pair 3.5kb, we generated, in a multiplex sequencing run 599,335.274 reads in
366 total with a Qscore greater than 18 (Table 4). When using the same classification approach as before
367 (wf-metagenomics workflow, Kraken2, ncbi_16s_18s_28s_ITS), we observed that over 99.9% of reads
368 were successfully classified for amplicons generated with all primer pairs. All classified reads were
369 assigned to fungi, which is attributed to the fact that the ncbi_16s_18s_28s_ITS database only contains
370 fungal sequences and lacks other eukaryotic sequences. Therefore, we reanalyzed the data using the
371 SILVA database, which includes a broader range of eukaryotes. For amplicons generated with primer
372 pair 7 97% of the reads were classified as fungi, with primer pair 9 95%, and with primer pair 10 97%.
373 Overall, none of the amplicons generated with the new primer pairs were classified as plant (Fig. S3).
374 For the amplicons generated with the non-discriminative 3.5 kb primer pair, however, only 13%-37%
375 of the reads were classified as fungi, while 60%-85% were classified as plant sequences. This
376 demonstrates that our newly-designed primer pairs are indeed able to efficiently discriminate between
377 the rDNA locus of fungi and the host plant. Moreover, when analyzed with the NCBI database, which
378 lacks plant entries, the plant reads were misclassified mostly as the fungal species *Brunneoclavispora*

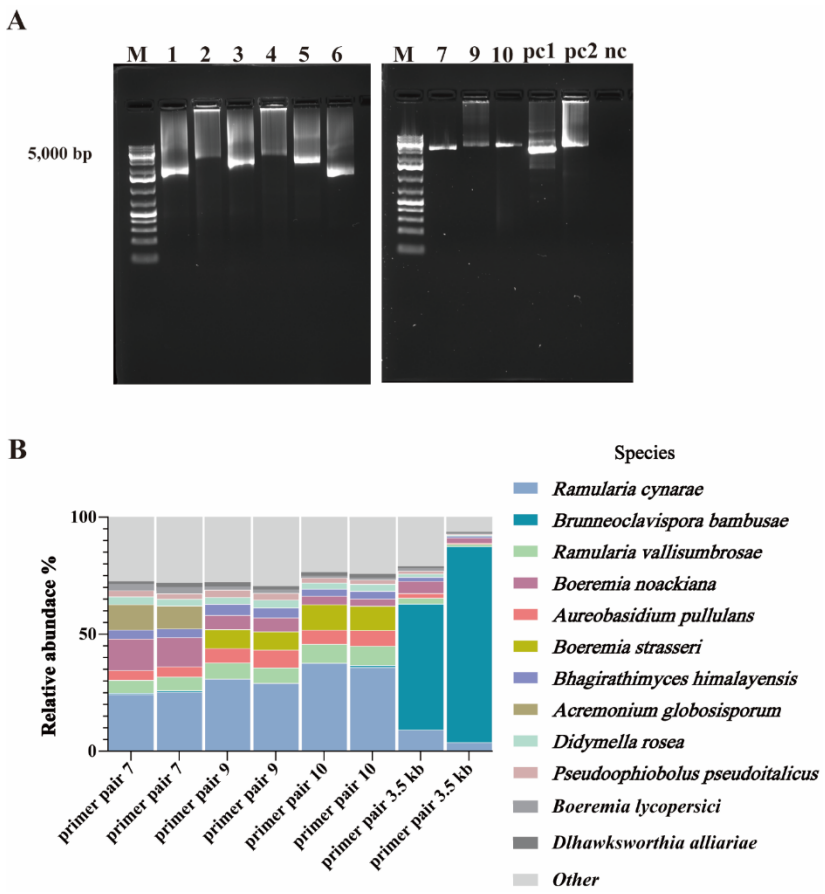

Figure 4. Metabarcoding of holobiont DNA of *V. minor*. (A) Gel electrophoresis of rDNA amplicons generated from gene specific primers including overhangs for PCR barcoding; Sample order in both panels: M: 1 kb plus DNA ladder; 1-10: amplicons generated with primer pairs 1-10 respectively; pc1: positive control generated with primer pair 3.5 kb; pc2: positive control generated with primer pair 10 with fungal mock community; nc: negative control (nuclease-free water). (B) Relative abundance of taxonomic groups classified based on the sequencing reads of amplicons generated from the genomic holobiont DNA of *V. minor* with the different primer pairs on species level (each primer pair has two replicates).

bambusae. This highlights that choosing the appropriate database for the research question at hand is crucial (Figure 4B).

Since the classification based on the ncbi_16s_18s_28s_ITS database appeared to be more robust for the fungal mock community, we performed additional analyses of the reads from the holobiont samples classified against this database. First, we performed a rarefaction analysis to investigate whether we reached an appropriate sequencing depth of our fungal community. We observed that the species richness asymptotically approaches a plateau with increasing number of reads, suggesting that the sequencing depth per sample was sufficient to estimate the microbial diversity (Figure 5A). Second, we analyzed the species diversity of the fungal community based on the amplicons generated with the

397 **Table 4** Read statistics after de-multiplexing of sequenced amplicons generated from the *V. minor*
398 holobiont genomic DNA with different primer pairs, and classification results..

Primer pair	Number of reads (after basecalling)	Number of reads (after trimming[a])	Mean length (bp)	Mean Qscore	Total yield (Mb)	Classification proportion (%)
7	61,766	49,377	3,056	18.4	151	99.90
	81,611	68,262	3,235	18.4	208	99.99
9	36,213	29,552	3,488	18.2	103	100.00
	72,599	55,396	3,499	18.2	194	99.99
10	74,663	64,301	3,682	18.2	237	99.94
	78,052	66,834	3,652	18.3	244	99.99
3.5 kb	76,702	63,370	2,745	18.2	174	99.99
	84,937	75,240	2,772	18.2	209	99.99
PC[b]	31,522	3,648	3,660	18.2	13.4	99.97
	22,274	2.979	3,653	18.2	10.9	100.00
NC[c]	1,248	21	2,734	16.3	5.74e-08	-

399 [a] Reads were trimmed in the range of 2kb – 6kb;
400 [b] Positive control with mock community as input along the whole experimental procedure;
401 [c] Negative control with only nuclease-free water as input along the whole experimental procedure.

402 different primer pairs. We noticed that the alpha diversity appeared to be greater when the amplicons
403 were generated with primer pairs 7 and 9 than with primer pairs 10 and 3.5 kb (Figure 5 B). This may
404 indicate that primer pairs 7 and 9 could potentially better capture the microbial diversity in the *V. minor*
405 mycobiome.

406 Discussion

407 Several approaches have been developed to address host contamination in the analysis of plant-
408 associated microbiota. One such method involves the use of PCR clamping with peptide nucleic acid
409 (PNA) to block the amplification of plant chloroplast 16S rRNA genes and mitochondrial 18S rRNA
410 genes (Lundberg *et al.*, 2013). However, this PNA-blocking technique has proven not very effective
411 and is associated with significant costs (Azadnia *et al.*, 2023). In this study, we developed fungus-
412 specific primers for the generation of long amplicons that almost entirely exclude plant DNA
413 demonstrating a cost-effective alternative to PNA usage.

414 High-throughput sequencing of taxonomic marker genes (metabarcoding) is a widely used and powerful
415 technique for biodiversity and microbial ecology. However, many taxa recovered from natural
416 environments remain unidentifiable at the species level or even higher taxonomic levels when working
417 with short reads (Tedersoo *et al.*, 2020). In this case, long-read sequencing has the potential to overcome
418 this problem. For instance, the identification of the six fungal isolates within our mock community based
419 solely on ITS sequences was inconclusive, as many isolates exhibited 100% sequence similarity with

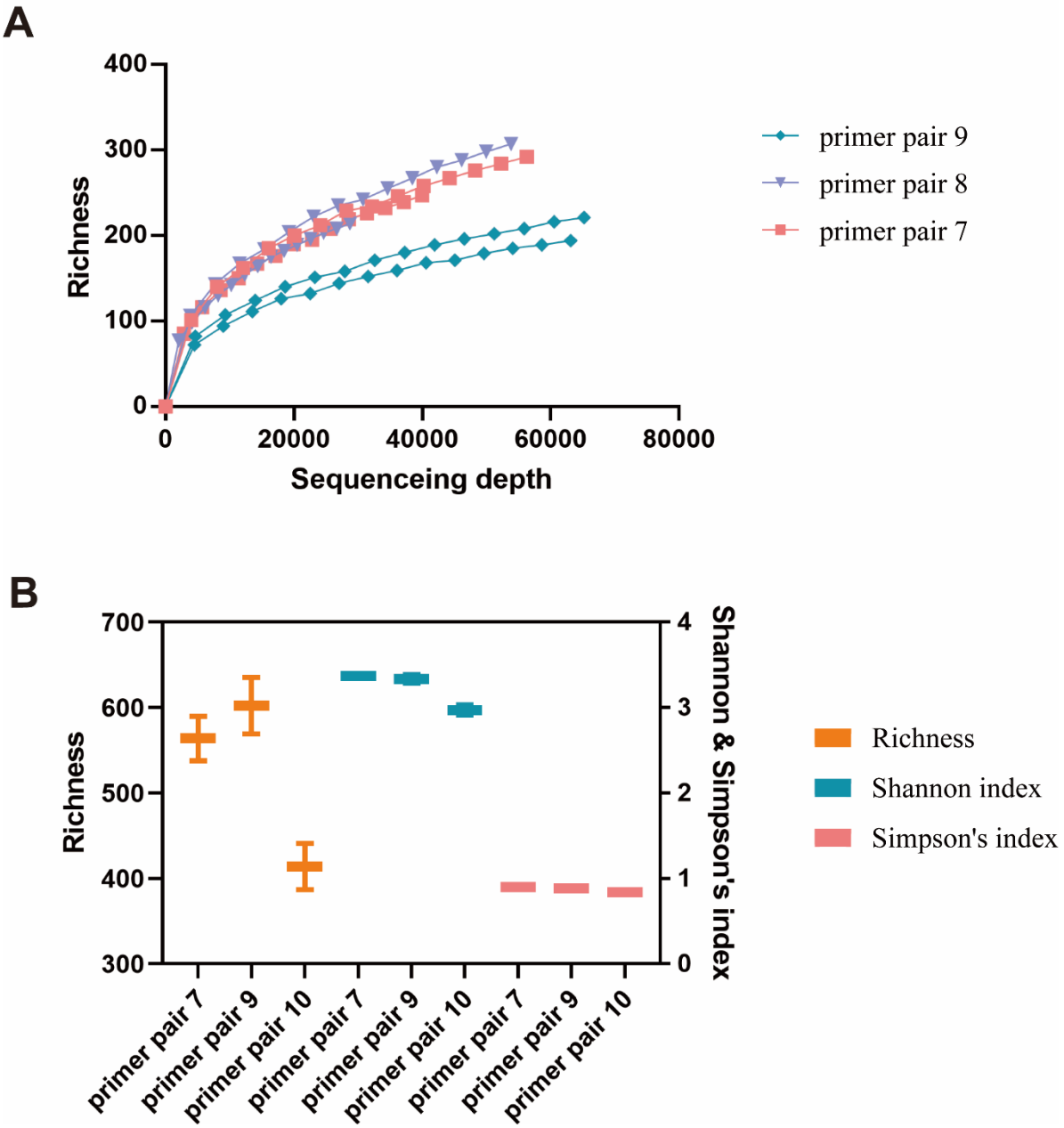

Figure 5. Alpha diversity analysis for the *V. minor* mycobiome. (A) Rarefaction curves; (B) Species richness, Shannon's index and Simpson's index.

various taxa, hindering precise taxonomic classification. For example, for the species *Thyridium oculorum* in the mock community, ITS-based identification returned matches to both *Phialemonium* sp. and *Thyridium* sp., thus preventing a definitive taxonomic assignment. However, long-read sequencing results from this study showed that 96.7% of all reads derived from this isolate were classified as *Thyridium oculorum*, which is consistent with the identification based on multi-locus classification. Similarly, the species assignments for *Trametes versicolor* and *Neocucurbitaria salicis-albae* were consistent with multi-locus classification and 100% and 99.15% of the reads derived from these isolates,

430 respectively, were correctly classified. The classification of *Cadophora luteo-olivacea* also agreed with
431 multi-locus identification, although only 75.7% of the reads were correctly classified. 23.7% of the reads
432 were classified as [*Graphium*] *rubrum*, an uncertain species from the same order. This highlights that
433 long-read sequencing can provide superior classification results in metabarcoding studies.

434 There are, however, still common pitfalls that need to be tackled in the future. First, reducing bias in the
435 relative abundance of sequence variants should be a primary objective in enhancing amplicon
436 sequencing methods. It is well established that even PCR primers designed for broad coverage tend to
437 amplify some phylogenetic subgroups more effectively than others. For instance in this study, the
438 inferred abundance of Thyridiaceae in the mock community was consistently lower than the expected
439 ratio, which might be attributed to amplification efficiencies of the used primer sets. Also in the
440 experiments with the holobiont DNA, we observed a striking difference in the formally most abundant
441 taxa across samples prepared with different primer pairs. This is an important aspect that needs further
442 consideration when choosing the appropriate primers for a study.

443 Second, the accuracy of taxonomic classification is highly dependent on the reference database
444 employed. Currently, these databases mainly rely on parts of the rDNA locus. The SILVA database
445 includes near full-length sequences of the small and large subunits of the rRNA operon for both
446 prokaryotes and eukaryotes. However, these sequences are insufficient for amplicons spanning the ITS
447 region and generally allow identification only up to the genus level. Despite this limitation, the SILVA
448 database enables the identification of non-fungal sequences, such as those from Chloroplastida and
449 Metazoa (Quast *et al.*, 2012). Conversely, for classification of fungi, the NCBI database seems to
450 outperform the SILVA database. This superior performance is due to NCBI's focus on the RefSeq
451 database, which includes ribosomal DNA loci (16S, 18S , 28S rDNA genes, and ITS region) and allows
452 for species-level identification, offering better accuracy and resolution in taxonomic classification of
453 microbes. The limitation of this database, however, is that it is focused on archaea, bacteria and fungi,
454 and does not include plants or other micro- and macro-eukaryotes. Therefore, reads originating from
455 such eukaryotes will either not be classified at all or incorrectly classified as fungi. Furthermore, the
456 fungal tree of life is still rather sparsely sampled for molecular barcodes deposited in the RefSeq
457 database. Thus, significant efforts are still needed to improve the database to enhance its completeness
458 and accuracy

459 Conclusions

460 In the current study, we have developed a host-exclusive primer design workflow, which is centered
461 around a new computation tool, mbc-prime, to design discriminative metabarcoding primers with the
462 option to exclude the DNA amplification of non-target taxa, for instance the host in host-associated
463 microbiomes. We demonstrated the use of mbc-prime when targeting endophytic fungi associated with

the medicinal plant *V. minor*. Initially, we evaluated our custom-designed primers *in silico*, demonstrating high coverage of fungal lineages with no coverage of Gentianales. Subsequent evaluation using a mock community revealed that the relative abundance of amplicons accurately reflected the community composition of the mock community. Moreover, when applied to plant holobiont samples, our primers successfully targeted a broad range of endophytic fungi while completely excluding the plant genome. This specificity has the potential to shed light on the roles of endophytic fungi in plant health and ecosystem dynamics. In addition, this demonstrates that our workflow, incorporating the mbc-prime tool, can effectively be used to study the composition of complex host-associated microbiomes. Furthermore, the primer design workflow developed in this study is universally applicable for designing discriminative primers targeting particular phyla, offering a user-friendly and practical solution for researchers in microbiome studies.

Acknowledgments

We thank the Center for Information Technology of the University of Groningen for their support and for providing access to the Peregrine and Hábrók high-performance computing cluster. KH is grateful for financial support from the Federation of Biochemical Societies through the FEBS excellence award. THe is supported by a scholarship 202006550001 from the China Scholarship Council.

Competing interests

No competing interest is to be declared.

Author contributions

THe, THa and KH designed the study and the workflow. THe carried out the primer design with contributions from KJ and BS. THe led the lab and field work with contributions from AR, XL and RS. THe carried out the benchmarking. THa developed and implemented the software with contributions from KJ. THe, THa and KH wrote the manuscript. THa and KH supervised the project.

Data Availibility

The mbc-prime tool is available on Github https://github.com/thackl/mbc-prime. All data will be made available upon publication of this manuscript in a peer reviewed journal or is available upon reasonable request from the corresponding author.

References

Abdelfattah A, Malacrinò A, Wisniewski M, Cacciola SO, Schena L. 2018. Metabarcoding: A powerful tool to investigate microbial communities and shape future plant protection strategies. *Biological Control* **120**: 1–10.

Azadnia A, Mikryukov V, Anslan S, Hagh-Doust N, Rahimlou S, Tamm H, Tedersoo L. 2023. Structure of plant-associated microeukaryotes in roots and leaves of aquatic and terrestrial plants revealed by blocking peptide-nucleic acid (PNA) amplification. *FEMS Microbiology Ecology* **99**.

Bahadori F, Topçu G, Boa M, Türkekul A, Kolak U, Kartal M. 2012. Indole alkaloids from *Vinca major* and *V. minor* growing in Turkey. *Natural Product Communications* **7**: 731–734.

Bolívar-Anillo HJ, Garrido C, Collado IG. 2020. Endophytic microorganisms for biocontrol of the phytopathogenic fungus *Botrytis cinerea*. *Phytochemistry Reviews* **19**: 721–740.

Ceballos-Escalera A, Richards J, Arias MB, Inward DJG, Vogler AP. 2022. Metabarcoding of insect-associated fungal communities: a comparison of internal transcribed spacer (ITS) and large-subunit (LSU) rRNA markers. *MycoKeys* **88**: 1–33.

Charnaud S, Munro JE, Semenec L, Mazhari R, Brewster J, Bourke C, Ruybal-Pesántez S, James R, Lautu-Gumal D, Karunajeewa H, *et al.* 2022. PacBio long-read amplicon sequencing enables scalable high-resolution population allele typing of the complex CYP2D6 locus. *Communications Biology* **5**: 168.

Cuscó A, Catozzi C, Viñes J, Sanchez A, Francino O. 2018. Microbiota profiling with long amplicons using Nanopore sequencing: full-length 16S rRNA gene and whole rrn operon. *F1000Research* **7**: 1755.

D'Andreano S, Cuscó A, Francino O. 2021. Rapid and real-time identification of fungi up to species level with long amplicon nanopore sequencing from clinical samples. *Biology Methods and Protocols* **6**: 1–6.

Dastogeer KMG, Tumpa FH, Sultana A, Akter MA, Chakraborty A. 2020. Plant microbiome–an account of the factors that shape community composition and diversity. *Current Plant Biology* **23**: 100161.

Fandy TE, Abdallah I, Khayat M, Colby DA, Hassan HE. 2016. In vitro characterization of transport and metabolism of the alkaloids: vincamine, vinpocetine and eburnamonine. *Cancer Chemotherapy and Pharmacology* **77**: 259–267.

Farahanikia B, Akbarzadeh T, Jahangirzadeh A, Yassa N, Ardekani MRS, Mirnezami T, Hadjiakhoondi A, Khanavi M. 2011. Phytochemical investigation of *Vinca minor* cultivated in Iran. *Iranian Journal of Pharmaceutical Research* **10**: 777–785.

De Filippis F, Laiola M, Blaiotta G, Ercolini D. **2017**. Different Amplicon Targets for Sequencing-Based Studies of Fungal Diversity (HL Drake, Ed.). *Applied and Environmental Microbiology* **83**: e00905-17.

Gardes M, Bruns TD. **1993**. ITS primers with enhanced specificity for basidiomycetes - application to the identification of mycorrhizae and rusts. *Molecular Ecology* **2**: 113–118.

Heeger F, Wurzbacher C, Bourne EC, Mazzoni CJ, Monaghan MT. **2019**. Combining the 5.8S and ITS2 to improve classification of fungi (D Yu, Ed.). *Methods in Ecology and Evolution* **10**: 1702–1711.

He T, Li X, Iacovelli R, Hackl T, Haslinger K. **2023**. Genomic and Metabolomic Analysis of the Endophytic Fungus *Fusarium* sp. VM-40 Isolated from the Medicinal Plant *Vinca minor*. *Journal of Fungi* **9**.

Hyde KD, Dong Y, Phookamsak R, Jeewon R, Bhat DJ, Jones EBG, Liu N-G, Abeywickrama PD, Mapook A, Wei D, *et al.* 2020. Fungal diversity notes 1151–1276: taxonomic and phylogenetic contributions on genera and species of fungal taxa. *Fungal Diversity* **100**: 5–277.

Ihrmark K, Bödeker ITM, Cruz-Martinez K, Friberg H, Kubartova A, Schenck J, Strid Y, Stenlid J, Brandström-Durling M, Clemmensen KE, *et al.* 2012. New primers to amplify the fungal ITS2 region - evaluation by 454-sequencing of artificial and natural communities. *FEMS Microbiology Ecology* **82**: 666–677.

Katoh K, Standley DM. **2013**. MAFFT multiple sequence alignment software version 7: Improvements in performance and usability. *Molecular Biology and Evolution* **30**: 772–780.

Klindworth A, Pruesse E, Schweer T, Peplies J, Quast C, Horn M, Glöckner FO. **2013**. Evaluation of general 16S ribosomal RNA gene PCR primers for classical and next-generation sequencing-based diversity studies. *Nucleic Acids Research* **41**: e1.

Lundberg DS, Yourstone S, Mieczkowski P, Jones CD, Dangl JL. **2013**. Practical innovations for high-throughput amplicon sequencing. *Nature Methods* **10**: 999–1002.

Monod V, Hofstetter V, Zufferey V, Viret O, Gindro K, Croll D. **2022**. Quantifying Trade-Offs in the Choice of Ribosomal Barcoding Markers for Fungal Amplicon Sequencing: a Case Study on the Grapevine Trunk Mycobiome (MZ Anderson, Ed.). *Microbiology Spectrum* **10**: e02513-22.

Nilsson RH, Anslan S, Bahram M, Wurzbacher C, Baldrian P, Tedersoo L. **2019**. Mycobiome diversity: high-throughput sequencing and identification of fungi. *Nature Reviews Microbiology* **17**: 95–109.

Parniske M. **2008**. Arbuscular mycorrhiza: the mother of plant root endosymbioses. *Nature Reviews Microbiology* **6**: 763–775.

Quast C, Pruesse E, Yilmaz P, Gerken J, Schweer T, Yarza P, Peplies J, Glöckner FO. **2012**. The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. *Nucleic Acids Research* **41**: D590–D596.

Raja HA, Miller AN, Pearce CJ, Oberlies NH. **2017**. Fungal Identification Using Molecular Tools: A Primer for the Natural Products Research Community. *Journal of Natural Products* **80**: 756–770.

Rodriguez RJ, Henson J, Van Volkenburgh E, Hoy M, Wright L, Beckwith F, Kim Y-O, Redman RS. **2008**. Stress tolerance in plants via habitat-adapted symbiosis. *The ISME Journal* **2**: 404–416.

Rosling A, Cox F, Cruz-Martinez K, Ihrmark K, Grelet G-A, Lindahl BD, Menkis A, James TY. **2011**. Archaeorhizomycetes: Unearthing an Ancient Class of Ubiquitous Soil Fungi. *Science* **333**: 876–879.

Rozen S, Skaletsky H. **2000**. Primer3 on the WWW for General Users and for Biologist Programmers. In: Bioinformatics Methods and Protocols. New Jersey: Humana Press, 365–386.

Schoch CL, Seifert KA, Huhndorf S, Robert V, Spouge JL, Levesque CA, Chen W, Bolchacova E, Voigt K, Crous PW, *et al.* 2012. Nuclear ribosomal internal transcribed spacer (ITS) region as a universal DNA barcode marker for Fungi. *Proceedings of the National Academy of Sciences of the United States of America* **109**: 6241–6246.

Šošić M, Šikić M. **2017**. Edlib: a C/C ++ library for fast, exact sequence alignment using edit distance (J Hancock, Ed.). *Bioinformatics* **33**: 1394–1395.

Tedersoo L, Albertsen M, Anslan S, Callahan B. **2021**. Perspectives and Benefits of High-Throughput Long-Read Sequencing in Microbial Ecology (IS Druzhinina, Ed.). *Applied and Environmental Microbiology* **87**: 1–19.

Tedersoo L, Anslan S, Bahram M, Kõljalg U, Abarenkov K. **2020**. Identifying the 'unidentified' fungi: a global-scale long-read third-generation sequencing approach. *Fungal Diversity* **103**: 273–293.

Tedersoo L, Anslan S, Bahram M, Põlme S, Riit T, Liiv I, Kõljalg U, Kisand V, Nilsson RH, Hildebrand F, *et al.* 2015. Shotgun metagenomes and multiple primer pair-barcode combinations of amplicons reveal biases in metabarcoding analyses of fungi. *MycoKeys* **10**: 1–43.

Tedersoo L, Bahram M, Põlme S, Kõljalg U, Yorou NS, Wijesundera R, Ruiz LV, Vasco-Palacios AM, Thu PQ, Suija A, *et al.* 2014. Global diversity and geography of soil fungi. *Science* **346**: 1256688.

Tedersoo L, Bahram M, Zinger L, Nilsson RH, Kennedy PG, Yang T, Anslan S, Mikryukov V. **2022**. Best practices in metabarcoding of fungi: From experimental design to results. *Molecular Ecology* **31**: 2769–2795.

586 **Tedersoo L, Jairus T, Horton BM, Abarenkov K, Suvi T, Saar I, Kõljalg U**. **2008**. Strong host
587 preference of ectomycorrhizal fungi in a Tasmanian wet sclerophyll forest as revealed by DNA
588 barcoding and taxon-specific primers. *New Phytologist* **180**: 479–490.

589 **Toju H, Tanabe AS, Yamamoto S, Sato H**. **2012**. High-Coverage ITS Primers for the DNA-Based
590 Identification of Ascomycetes and Basidiomycetes in Environmental Samples. *PLoS ONE* **7**: e40863.

591 **Vas Á, Gulyás B**. **2005**. Eburnamine derivatives and the brain. *Medicinal Research Reviews* **25**: 737–
592 757.

593 **Vu D, Groenewald M, de Vries M, Gehrmann T, Stielow B, Eberhardt U, Al-Hatmi A,**
594 **Groenewald JZ, Cardinali G, Houbraken J, *et al.* 2019**. Large-scale generation and analysis of
595 filamentous fungal DNA barcodes boosts coverage for kingdom fungi and reveals thresholds for fungal
596 species and higher taxon delimitation. *Studies in Mycology* **92**: 135–154.

597 **White TJ, Bruns T, Lee S, Taylor J**. **1990**. Amplification and direct sequencing of fungal ribosomal
598 RNA genes for phylogenetics. In: PCR Protocols. Elsevier, 315–322.

599 **Wurzbacher C, Larsson E, Bengtsson-Palme J, Van den Wyngaert S, Svantesson S, Kristiansson**
600 **E, Kagami M, Nilsson RH**. **2019**. Introducing ribosomal tandem repeat barcoding for fungi. *Molecular*
601 *Ecology Resources* **19**: 118–127.

602 **Yap M, Feehily C, Walsh CJ, Fenelon M, Murphy EF, McAuliffe FM, van Sinderen D, O'Toole**
603 **PW, O'Sullivan O, Cotter PD**. **2020**. Evaluation of methods for the reduction of contaminating host
604 reads when performing shotgun metagenomic sequencing of the milk microbiome. *Scientific Reports*
605 **10**.

606 **Ye J, Coulouris G, Zaretskaya I, Cutcutache I, Rozen S, Madden TL**. **2012**. Primer-BLAST: A tool
607 to design target-specific primers for polymerase chain reaction. *BMC Bioinformatics* **13**: 134.

608

609 Supporting Information

610 Word document

611 **Fig. S1.** The design of degenerate primers illustrated in the alignment of fungi, plants, and primer
612 sequences.

613 **Fig. S2**. Barplot of the 9 most abundant taxa at the genus rank in all fungal community samples using
614 the SILVA database.

615 **Fig. S3**. Barplot of the 12 most abundant taxa at the genus rank in all plant holobiont samples using
616 SILVA database.

617 **Table S1**. Metadata for different primer combinations.

618 **Table S2.** Fungal isolates used and their contribution to the mock community by pooling of rDNA at
619 equimolar concentrations.

620 **Notes S1**. A step-by-step tutorial for using mbc-prime.

621 Excel document

622 **Dataset S1.** 118 fungal SSU sequences from different phyla downloaded from the SILVA database.

623 **Dataset S2.** 46 sequences from the order Gentianales downloaded from the SILVA database.

624 **Dataset S3.** 62 fungal sequences downloaded from the SILVA database.

625 **Dataset S4.** 6 sequences from Gentianales downloaded from the SILVA database.

626 **Dataset S5.** 147 candidate primers obtained from mbc-prime across the SSU region.

627 **Dataset S6.** 371 candidate primers obtained from mbc-prime across the LSU region.

628 **Dataset S7.** Optimized primer sequences used in this study.

629 **Dataset S8.** The metadata of samples as input for the library preparation and sequencing.